

School of Informatics
University of Edinburgh



Thesis Proposal

Ontology Evolution in Physics

by

Michael Chan

December, 2008

Primary Supervisor:
Secondary Supervisor:

Prof. Alan Bundy
Dr. Jos Lehmann

Contents

1	Introduction	3
2	Literature Survey	5
2.1	Agents and Ontologies	5
2.2	Ontology Dynamics	6
2.3	The Ontology Repair System	7
2.4	Ontology Languages	8
2.5	Non-monotonic Reasoning	9
2.6	Belief Revision	10
2.7	Scientific & Mathematical Discovery	12
3	Proposal	13
3.1	Development of Ontology Repair Plans	14
3.1.1	The “Where’s my stuff?” Ontology Repair Plan	14
3.1.2	The Inconstancy Ontology Repair Plan	16
3.2	Analysis of Case Studies and Development of Ontologies	17
3.2.1	Discovery of Latent Heat	17
3.2.2	Discovery of Dark Matter	17
3.2.3	Modified Newtonian Dynamics	18
3.2.4	Measurement of the Travel Time of Light	19
3.2.5	Discovery of Atmospheric Pressure	20
3.3	Implementation	22
3.4	Evaluation of Repaired Ontologies	23
3.4.1	Application of WMS to the Discovery of Latent Heat	23
3.4.2	Application of WMS to Dark Matter	24
3.4.3	Application of Inconstancy to Modified Newtonian Dynamics	25
3.4.4	Application of Inconstancy to the Travel Time of Light	26
3.5	Contribution to Knowledge	26
4	Future Direction	27
4.1	Overview	27
4.2	Development of a Theory for Ontology Evolution	28
5	Programme of Work	30
5.1	Potential Risks	32
6	Conclusion	33

Abstract

This thesis proposal outlines a research project on automated ontology evolution in the physics domain. We argue that intelligent agents must be able to manipulate their own ontologies autonomously in order to communicate effectively in an open, heterogeneous environment. Patterns of fault diagnosis and resolution in reasoning, based on human problem solving skills in physics, are compiled together to form plans executable by agents. The physics domain has the advantage that many faulty ontologies, and the reasoning required to mend them, have been well-documented.

1 Introduction

Epistemology, the study of knowledge, has a long history in philosophy. Initiated by Descartes and Locke, the work during the early era was based on intuitive, rationalistic and introspective modes of reasoning. To eliminate subjectivism, Leibniz proposed a need for a rigorous formalisation of reasoning that would allow errors to be as easily detected as in arithmetic. Leibniz's idea of enhancing the role of mathematics and using a logical structure of reasoning evolved to become the concept of a universal scientific language, known as *Characteristica Universalis*. Other concepts, such as Kant's system of categories, have also demonstrated the power of formal representation and logical analysis of reasoning.

The perspective of a formal approach has been widely accepted and supported by the contemporary scientific community at large. Such a rigorous approach can provide analysis of reasoning with precision and reliability; consequently, it can benefit all areas concerned with the role and state of knowledge. Motivated by various applications in the past couple of decades, significantly more attention has been drawn from researchers in various fields ranging from philosophy, economics, and linguistics to theoretical computer science and artificial intelligence. The research focuses across these fields are generally different, but it is typically conceded that appropriate representation of knowledge is crucial for effective reasoning, as argued by Pólya [50].

AI and, more generally, computer science are presently faced with the challenge that intelligent agents must be able to represent and manipulate their own knowledge. For an agent to perform reasoning, the conceptualisation of the entities of an application domain is usually represented in the agent's *knowledge-base*. Unfortunately, the world is not static, so changes in our conceptualisation of this infinitely-complex domain are necessary [7]. Knowledge-base evolution, the process of updating a knowledge-base when new information is acquired, poses significant challenges to the representation of knowledge and the formalisation of reasoning by, for example, introducing inconsistency, ambiguity, and incompleteness into the knowledge-base. There are several types of knowledge-base evolution, including database schema evolution and, the focus of the proposed project, ontology evolution. Defined by Stojanovic et al. in [60], ontology evolution is "the timely adaptation of an ontology to the arisen changes and the consistent propagation of these changes to dependent artefacts". Although Stojanovic et al.'s focus is on user-driven approaches, the definition is also appropriate for automated ontology evolution.

The need to reliably handle ontology evolution has been intensified by the increased demands created by multi-agent systems, for example, the vision of

agents interacting in the Semantic Web. Investigations in ontology evolution will elucidate new mechanisms for more accurate agent communication and give a better understanding of the general knowledge-base evolution process. If autonomous systems are to deal with the ever-changing world, they must be able to autonomously update their own ontologies. This requires them to be able to both detect faults in and manipulate their theory of the world. The failure to update the theory accordingly will usually lead to ineffective or inconsistent communication, caused by the discrepancy between different conceptualisations of the world. Such updates must go beyond the ability to change beliefs and learn new concepts in terms of the old ones, and in these cases the underlying syntax and semantics of the ontology themselves may also require alterations. For instance, if sales tax is to be introduced into an inventory system, then the inventory agent could keep the original signature and include sales tax in the calculation of an item's price. However, it would be more appropriate for the inventory agent to increase the arity and give the predicate representing the item price an additional argument to indicate the corresponding amount of sales tax. Interactive manipulation is only useful at design-time and scales up to a certain complexity, so mechanisms for *automated* ontology evolution are highly desirable.

As described in §2.2, there are many types of ontology change: ontology evolution, mapping, morphism, and matching. Ontology evolution is considered to be the most radical form as it manipulates the profound structure of the ontology [13]. As opposed to the other types, ontology evolution implements changes to both the syntax and the semantics of the ontology. Automated ontology evolution explores how the theory should evolve given some new conflicting information and how an agent adapts to the changing world and goals. Typically, the original theory itself is consistent, but the merge with the newly obtained information leads to inconsistency. This process may be seen as analogous to biological evolution because the ontology automatically adapts to its changing environment. For instance, if the observed value of a function unexpectedly varies when the theory predicts it to remain constant, then the theory should be mended to consider the parameter causing the variation as an argument of the function. There are many other similar reasoning faults that may occur in problem solving, and they are eventually resolved by the solver's intelligence and creativity. Hence, we believe human reasoners' imagination used in problem solving can help advance mechanisms for ontology evolution.

In the proposed project, we will explore a) patterns for detecting ontological contradictions arising from taking into account some new information, and b) heuristics for resolving the resulting conflicts between concepts – we call the composition of such patterns *ontology repair plans*. Repair plans are generic combinations of diagnosis and repair operations that guide the ontology evolution process. To identify such patterns, we investigate records of ontology evolution in the history of physics. Some of the most seminal advances in the development of physics required ontology evolution. Moreover, the evolutionary process in physics is generally very well documented. Detailed accounts are available of: the problems with the prior ontology, e.g., a contradiction between theory and experiment; the new, consistent ontology; and an account of the reasoning which lead to it.

Many physics properties are better represented as higher-order objects, so the repair plans are to be implemented in a higher-order, polymorphic, logic

programming language. For instance, a star’s orbit is more realistically represented as a function object that returns the 3-D spatial position of the star for a given time moment. If it is represented as a non-functional object, then much of the description and expressivity will be lost and the representation will be similar to that of static, physical entities, e.g., a dog or a house. Moreover, polymorphism is crucial to the generality of the implementation, e.g., comparison operators such as $=$, $<$ and $>$ need to apply to a variety of types. The implementation of ontology repair plans and example ontologies will be part of the *Guided Analyses of Logical Inconsistency Lead to Evolved Ontologies* (GALILEO) system.

The aim of this project is to demonstrate that automatic ontology evolution via repair plans is computationally feasible and can account for the kinds of ontology evolution that are observed in human problem solving in physics. The specific hypothesis that will be evaluated in the project is that:

A few generic, ontology repair plans can account for a large number of historical instances of ontology evolution in the physics domain.

The evaluation of the repair plans will assess to what extent they create a new ontology that escapes the failures diagnosed in the prior ontology and to what extent this emulates the historical process of ontology evolution.

In the rest of the proposal, §2 briefly describes existing literature relevant to the proposed project. §3 presents an outline of the research in more depth and summarises work done to date. This includes a breakdown of the methodology into manageable components. §4 outlines possible avenues for future research. §5 presents a work-plan, approximating the effort required to complete each major component.

2 Literature Survey

Although relatively little research effort has been directed towards automated ontology evolution, this section briefly outlines a rich literature related to the aim and the general methodology. §2.1 presents the background behind agent technology and the concept of ontology. A major problem arising from the communication of agents in an open environment is ontological heterogeneity; various techniques used for seeking agreement between semantically incompatible ontologies are described in §2.2. Of the many proposals for mechanisms to resolve ontological mismatches, §2.3 outlines the *Ontology Repair System* in more details. In recent years, greater consensus on the representation of ontologies has been reached; §2.4 presents an overview of several popular meta-representation languages and a discussion on their semantics. For a review of prominent ideas dealing with the related semantics and reasoning, §2.5 focuses on non-monotonic reasoning and §2.6 describes work done in belief revision. Finally, §2.7 provides an idea of (semi-) automated scientific reasoning by summarising several successful scientific and theorem discovery systems.

2.1 Agents and Ontologies

Potential problems that arise from communication among heterogeneous agents in an open environment have stimulated studies on appropriate software paradigms.

The most widely recognised and accepted paradigm is the implementation of multiagent systems, or simply *agents* [66]. Interest in the field has been growing rapidly over the last decade, due to the increased emergence of applications involving vast distributed systems. For example, the realisation of the Semantic Web vision relies heavily on the success of agent technology [24, 58]. Depending on the type of problem being confronted, an agent should satisfy a set of properties that define its basic characteristics. For many researchers, *autonomy* is a definitional prerequisite and is useful in distinguishing agents from other types of intelligent software. This provides agents with the ability to “act without direct intervention from humans and have some control over the agent’s own actions and internal states” [8]. Agents should also exhibit several other characteristics, including *social ability*, *reactivity*, and *proactivity*. By these, an agent is expected to be able to interact with other entities, perceive and analyse its environment, and exhibit goal-directed behaviour by taking the initiative [65].

The introduction of agents has spurred research into ontologies: into their representation, engineering and use in reasoning. To interact with the environment, agents rely on ontologies which allow them to function by using possible concepts and the relationships among them. An ontology usually represents a certain view of a particular domain, in which the concepts are given unambiguous and explicit representations. This has led to a definition of ontology commonly adopted by the knowledge-engineering community. An ontology is regarded as “an explicit specification of a conceptualisation” [21], which is based on the formal notion of conceptualisation introduced by Genesereth and Nilsson in [19] and actually goes back to Tarski.

The ontologies used in this project are formulated as generic logical theories, not restricted to ontologies in Description Logics, KIF or other logics in which ontologies have been traditionally formalised. As logical theories, ontologies here consist of two key parts: the *signature*, which declares the functions and their types, and the *theory*, which is the set of theorems, defined recursively via the axioms and the rules of inference over the signature.

Finally, there are various types of ontologies, depending on the ontology’s level of abstraction. Defined in philosophy as a theory of existence, *top-level* ontologies typically contain fundamental knowledge and describe the most general concepts that are assumed to form common-sense. Hayes’ ontology for naive physics [23] is one of the earliest works that concentrate on the formalisation of the ontological features of the world. Other prominent examples of top-level ontologies include CYC’s ontology [33], DOLCE [16], Sowa’s top-level ontology [59], and SUMO [43]. Next in line are core ontologies covering fields of practice, e.g., LKIF-Core for law [25] and UMLS for medicine [4]. Domain ontologies contain more specialised knowledge, e.g., Gene Ontology [2] for describing genes.

2.2 Ontology Dynamics

As the world inevitably evolves over time, it is unrealistic to assume static knowledge and information. A common source of errors in knowledge-bases is due to changes in the application domain and user needs [61]. This often arises in database systems, where much work has been done to allow the schema to evolve in databases [37, 54, 63]. Causes of ontological heterogeneity include changes in the perspective [46] or in the conceptualisation [49] of the world. To address

these issues, automated repair mechanisms to resolve logical inconsistencies and to improve repair efficiency are required.

There are various types of approaches to ontological heterogeneity resolution. One type is to relate the signature of one ontology to another that shares the same domain of discourse by means of a function – this is known as *ontology mapping* [26]. It is somewhat different from *ontology morphism* [26, 35], which also includes mappings between axioms. A relatively more popular approach for ensuring semantic agreement between ontologies is *ontology matching*. Most matching techniques compute semantic distances between concepts in order to discover relationships between the signatures belonging to the ontologies. WordNet, machine learning and probabilistic techniques are popular methods to improve the quality of the output alignment. For instance, [11] exploits information in the taxonomical structure of the ontology and utilises a probabilistic model to combine the results of a set of learners to find relationships. For more complex matching, [20] grounds the concepts in an ontology in WordNet terms and formulates the task to become a constraint-satisfaction problem.

Interactive *ontology evolution*, which requires users to explicitly guide the evolution process, constitutes most research activity in ontology evolution. In [60], a six-phase evolution cyclic process is introduced, which systematically verifies the consequences of the major operations: *change capturing*, *change representation*, *semantics of change*, *change implementation* and *change propagation*. The implementation has been made within the KAON framework [5]. There is also related work on repairing inconsistencies in OWL ontologies: for instance [27], which focuses on strategies for removing inconsistent axioms and for identifying syntactical modelling errors in OWL ontologies to assist users to rewrite faulty axioms. [22] addresses ontology evolution by presenting algorithms for debugging inconsistent ontologies. To maintain consistency in the resulting ontology, some axioms in the original are removed. As pointed out in [18], the operations adopted to keep the resulting ontology consistent in fact resemble those in belief revision and contraction.

While extensive work has been done in handling ontological heterogeneity, automated ontology evolution uncovers a new area of research problems. The general goal of alignment techniques, e.g., ontology mapping, morphism, and matching, is to find semantic relationships between the ontological concepts defined in the signatures of two different ontologies. In automated ontology evolution, the main goal is not to find relationships between signatures automatically, but to make persistent changes to the signature itself. A major challenge is to investigate the range of possible conflicts arising from the heterogeneity and the ontological changes required to resolve the inconsistency.

2.3 The Ontology Repair System

The Ontology Repair System (ORS) [41] addresses the problem of ontology alignment in a multi-agent planning environment: a problem which must be solved in order to realise the vision of the Semantic Web. In the ORS environment, some agents offer services and others require these services. Each agent represents these services with STRIPS-like planning action rules with preconditions and effects written in a restricted version of KIF. It is inevitable with any sufficiently large agent community that there will be differences between their ontologies, even when there have been attempts to standardise them. The purpose of ORS

is to identify and repair these ontological mismatches at run time, since compile-time alignment is unrealistic in this scenario. Moreover, this had to be done without full access to the other agents' ontologies, since full access is also unrealistic. ORS's ontology *repair* differed from the more common ontology *matching* in being aimed at identifying and correcting errors in a single ontology rather than constructing a mapping between two ontologies. Ontology repair is also done automatically, dynamically and without access to the ontologies of other agents, whereas typically ontology matching has some manual element, is done statically and with full access to both ontologies.

The ORS ontology repair operations consist mostly of syntactic manipulations of the underlying logical representation. For instance, (a) the number or order of the arguments of a function may be changed; (b) a function may be divided into two or more, or two or more may be merged into one.

Such manipulations are similar to those envisaged for our work; however, we also expect to investigate more complex conceptual faults commonly occurred in problem solving.

2.4 Ontology Languages

Another aspect of ontology research and development is the language used for authoring ontologies. The ontology language sets the degree of expressiveness and tractability of the representation. Although the overall development is still at its early stage, ontology languages have significantly evolved to provide better support to ontology integration and interoperability.

Languages based on a markup scheme, e.g., XML, have been recommended as standard languages for the Semantic Web. Since the beginning, the standard language has been constantly evolving – from RDF to OWL. The *Resource Description Framework* (RDF) [36] is extended from XML, such that relationships between resources, i.e. subjects, predicates and objects, are expressed as triples. Such representation allows easier extraction of the subjects and objects in question, without the need to refer to a schema. However, RDF does not offer mechanisms for describing the attributes of or the relationships between resources. RDF Schema (RDFS) [6], a semantic extension of RDF, is a language for describing RDF vocabulary and structuring resources in an object-oriented manner. RDFS is still limited in various aspects, e.g., it cannot express useful constraints on predicates, such as those concerning cardinality and existence. That said, one cannot express that prime numbers are natural numbers that have exactly *two* divisors. The *Web Ontology Language* (OWL) [40] supports such constraints, uses RDF syntax, and with three variants, two of which are based upon description logics [3].

Description Logic (DL) is a knowledge representation paradigm for representing concepts and their hierarchies. Classes of instances are called *concepts*, which are interpreted as sets. Properties and relationships between concepts are represented as binary predicates called *roles*, e.g. *has_child(x, y)*. DL knowledge-bases comprise two components: TBox, which introduces terminology and defines the vocabulary, and ABox, which contains assertions over the terms in the TBox. DL implements *open-world semantics*, under which, as long as a fact is not proven to be inconsistent with the knowledge base, it is assumed to be unknown. Conventional relational databases employ the *closed-world semantics*, which assumes the knowledge-base contains all relevant facts and if a fact

cannot be derived, it is assumed to be false. In contrast, it makes more sense to use open-world semantics for representing ontologies in the Semantic Web, due to the enormous and partially capturable World Wide Web. A deductive reasoner for the Web, therefore, should not assume a statement to be true or false on the basis of a failure to *prove* or *disprove* it.

2.5 Non-monotonic Reasoning

Incomplete information prevails in our everyday reasoning as it is unrealistic to assume that conclusions can be drawn based on *all* relevant information. Further, it is unreasonable to assume that knowledge is static and that updates are not required. Classical logic systems typically are monotonic, i.e. if a sentence p follows from a set of propositions A , then p also follows from a set B , where $A \subset B$. Monotonicity does not allow the system to make inferences which may later be retracted when further information is added. Given two premises “birds fly” and “Tweety is a bird”, an agent can infer that “Tweety flies”. In the light of additional information that “Tweety is a kiwi”, monotonic systems cannot retract theorems even when they should.

The closed-world assumption (CWA), proposed by Reiter [52], is the earliest well known non-monotonic reasoning scheme. The idea is to assume all facts not specified to be false, so a system employing the CWA does not require explicit representations of negative facts in order to derive negative inferences. Being the basis of database theory, a flight database, for instance, only needs to contain details of all known flights to answer queries concerning both existing and non-existing flights. Clearly, the expressivity of CWA is very limited, as the underlying assumption is very strict, in the sense that all non-provable propositions are assumed false.

A more flexible formalism is Reiter’s *Default Logic* [51], which allows variables to be assigned values under “normal circumstances”. Default statements, or defaults, are treated as inference rules, rather than formulae. Defaults are in the form $\frac{\varphi:M\psi}{\psi}$, which captures the intuition that if the *prerequisite*, φ , holds and the *justification*, $M\psi$, can be consistently assumed, then the *conclusion*, ψ , can be inferred. In the Tweety example, $M\psi$ is inconsistent with what is known because kiwis cannot fly. The set of defaults for the Tweety example may contain, e.g., the default $\frac{bird(x):Mfly(x)}{fly(x)}$. Defaults can be paired with a set of known facts containing, e.g., $bird(tweety)$ for expressing that Tweety is a bird, to become a *default theory*.

McCarthy’s *circumscription* approach is another way of formalising non-monotonic reasoning; note that it is not a non-monotonic logic but a non-monotonic reasoning based upon first-order logic [38]. The underlying motivation is to let the tuples that satisfy the predicate be only those that follow from the sentence. From a model-theoretic viewpoint, the sentences yielded are true in all those models with the least number of tuples satisfied. This is a general formalisation of the CWA, as what is not specified is assumed false. Circumscription and Default Logic differ in several ways: for example, circumscription works with minimal models while Default Logic works with arbitrary models. McCarthy also introduced the concept of an abnormality predicate, denoted as $ab(x)$ [39], which can be used to represent that the proposition does not hold in normal circumstances. For instance, for the Tweety example, one

could express that birds normally fly with $\forall x. \text{bird}(x) \wedge \neg \text{ab}(x) \rightarrow \text{fly}(x)$ [53].

2.6 Belief Revision

An example non-monotonic form of reasoning is the process of changing an agent's beliefs to accommodate new information, possibly inconsistent with existing beliefs. The major focus of research here is to investigate possible models of belief change, known as belief revision operators, and demonstrate that they exhibit properties that resemble intuitive rationality. Rational belief revision operators must adopt reasonable, coherent revision.

Although there are many variations to the definition of rationality, almost all of them incorporate the principle of *minimal change*, which governs the need to preserve as much of earlier beliefs as possible. Unfortunately both in philosophy and in artificial intelligence, there is no single answer to achieving minimal change. The best known attempt to characterise minimal change is perhaps the AGM model [1], which has considered three forms of belief change: expansion, for adding a new belief to the belief set without regard to consistency; revision, for adding a new belief to the belief set and removing beliefs to maintain consistency; and contraction, for removing a belief from the belief set. The postulates for a single-round of revision, which is the most trivial belief revision strategy, are as follows:

- (R1) $K * \alpha$ is a belief set. *Revising K with α gives a belief set.*
- (R2) $\alpha \in K * \alpha$. *Revising K with α gives a set containing α .*
- (R3) $K * \alpha \subseteq \text{Closure}(K \cup \{\alpha\})$. *The belief set revised with a new belief contains only beliefs implied by the combinations of the old beliefs with the new belief.*
- (R4) If $\neg\alpha \notin K$, then $\text{Closure}(K \cup \{\alpha\}) \subseteq K * \alpha$. *If the new belief is consistent, then the beliefs implied by the combination of the old beliefs and the new belief make up the revised belief set.*
- (R5) $K * \alpha = \text{Closure}(\text{false})$ if and only if $\vdash \neg\alpha$. *The revised belief set is inconsistent if and only if the new belief is inconsistent.*
- (R6) If $\alpha \leftrightarrow \beta$, then $K * \alpha = K * \beta$. *The revised revision process abides by the principle of Irrelevance of Syntax, i.e. not to be affected by the syntactical forms of the new belief.*
- (R7) $K * (\alpha \wedge \beta) \subseteq \text{Closure}((K * \alpha) \cup \{\beta\})$. *The belief set revised with $\alpha \wedge \beta$ contains only beliefs implied by the combination of the beliefs revised with α and β .*
- (R8) If $\neg\beta \notin K * \alpha$ then $\text{Closure}((K * \alpha) \cup \{\beta\}) \subseteq K * (\alpha \wedge \beta)$. *If β is consistent with the belief set revised with α , then the beliefs implied by the combination of the beliefs revised with α .*

where K is a belief set and α and β are formulas. In [28], Katsuno and Mendelzon revised the AGM postulates to restrict a state of belief to a propositional formula. A revision operator $*$ on knowledge sets (theories) satisfies R1 to R6 iff a corresponding revision operator \circ on beliefs satisfies KM1 to KM4:

- (KM1) $\psi \circ \mu$ implies μ [\equiv R2]
- (KM2) If $\psi \wedge \mu$ is satisfiable, then $\psi \circ \mu \equiv \psi \wedge \mu$ [\equiv R3 and R4]
- (KM3) If μ is satisfiable, then $\psi \circ \mu$ is also satisfiable [\equiv R5]
- (KM4) If $\psi_1 \equiv \psi_2$ and $\mu_1 \equiv \mu_2$, then $\psi_1 \circ \mu_1 \equiv \psi_2 \circ \mu_2$ [\equiv R6],

where ψ is a propositional formula representing a knowledge-base and μ is a sentence respectively.

The semantic model of the AGM postulates is a partial pre-order of interpretations for expressing the plausibility, or the *epistemic entrenchment*, of interpretations [17]. The interpretation with the highest degree of epistemic entrenchment is given the highest rank and is designed to be the most preferred. Ideally, the most epistemically entrenched interpretation is the one that is the most appropriate after revising the knowledge-base with the new information, while maintaining rationality in the process. However, there is no unique solution to the assignment of ranks.

Both the notation and the semantics of the AGM theory have been criticised. Friedman and Halpern [15] consider the functional form of $*$ as problematic, as it suggests that the agent would have an ordering on epistemic entrenchments even when the epistemic state is given as a collection of formulae. Languages such as propositional and first-order logics would not be expressive enough to describe the ordering. If the entrenchment information is implicit in the revision operator $*$, then we cannot assume the same $*$ will be used in each iteration when performing iterated revision, since the relative entrenchment of beliefs is not necessarily maintained after each iteration.

Since the AGM theory is designed for mere one-step revision, it is too weak and not adept for handling *iterated revision*, which is the process of revising beliefs with new information iteratively. Iterative belief revision is closely connected to *conditional beliefs*, i.e. beliefs depending on earlier beliefs. Darwiche and Pearl [10] have proposed additional postulates which are designed for preserving the coherence of both conditional and unconditional beliefs, and for achieving absolute minimal change in conditional beliefs. Similarly, Freud and Lehmann [14] have described an additional postulate ensuring that if the new information is inconsistent, none of the previous beliefs is to be retained. Many other postulates have been proposed, but all behave well only in a specific problem setting.

None of the work in this area considers syntactical changes, so the representation of the world is assumed to be static. This assumption contradicts that underpinning ontology evolution, because the conceptualisation of the world is expected to change; therefore, the signature of an ontology is also expected to be updated. Furthermore, although the AGM postulates formalise the set-theoretic properties of the revision process, belief revision techniques are typically studied by assuming the underlying language is propositional logic. In contrast, the complex descriptions of physics objects and properties will require a much more expressive language, e.g., higher-order logic. However, the philosophical issues that belief revision confronts are somewhat similar to those faced by ontology evolution, e.g., rationality and minimal change. To develop a theory of ontology evolution, we might want postulates in a similar style as the AGM.

2.7 Scientific & Mathematical Discovery

As widely agreed by AI researchers, intelligence involves creativity. Progress towards more intelligent machines has motivated attempts at tackling problems known to be seemingly solvable only by human creativity [55]. A prime example problem is scientific and mathematical discovery, which requires the researcher to apply creative reasoning in order to invent new theories [31]. Due to the enormous search space, most work in this area focuses on formalising human reasoning abilities using heuristics.

One of the earliest heuristic-based machine discovery systems is Lenat’s Automated Mathematician (AM) program. It discovers mathematical concepts by performing concept formation and conjecture making in number theory [32]. It starts with about 110 elementary concepts, where each concept is represented by a set of slots. Each slot contains information regarding, e.g., definition, examples, generalisations/specialisations and worth. To fill these initially blank slots, AM looks through a database containing about 250 heuristics. Four types of heuristics were used: fill rules, for filling the slots; check rules, for validating the entries; suggest rules, for generating new concepts and new tasks; and interest rules, for measuring the interestingness of the concept. It is important to note that AM is designed to be interactive and the user has a crucial role in shaping the search process. The user can, for example, judge the interestingness of the theorems formulated so far. Some interesting results include the discovery of addition, multiplication, primes, and Goldbach’s conjecture.

Another system guided by heuristics is found in the BACON series of programs [30]. These programs are purely data-driven, i.e. generate theories from empirical data only. The heuristics help formulate regularities, such as constancies, trends, common divisors, and constant differences, hidden in the input data. One heuristic is that if the absolute values of a variable X increases as the absolute values of a variable Y increases without these values being linearly related, then their ratio should be considered. This heuristic was used to deduce Kepler’s laws as the data about planetary coordinates and trajectories show that the distance D increased with the period P , but the two are not linearly related; thus to get the $\frac{D}{P}$ form. These heuristics, therefore, can be seen as *ad-hoc* patterns for curve-fitting.

Pease, Smaill, Colton & Lee have used ideas from Lakatos to evolve mathematical theories [48]. For instance, they have automated the repair of faulty conjectures using computational versions of Lakatos’s *Proofs and Refutations* methods [29]. The proposed methods for fixing faulty conjectures in the light of counterexamples include: *monster-barring*, for modifying the definition to exclude unwanted counterexamples; *piecemeal exclusion*, for restricting the conjecture to those examples that do not exhibit properties of the counterexample; and *strategic withdrawal*, for restricting the conjecture to those examples for which it is known to hold. Their approach has contrasted with that proposed here by focusing on adding and changing definitions and conjectures within a signature that is changed only by definitional extensions. Our work also employs definitional extension, but is principally focused on more radical signature changes.

The proposed GALILEO system is also guided by heuristics, which formalise the methodology a problem solver adopts to reduce the search space. Both GALILEO and BACON target at the physics domain, but the two systems are

strategically different. For example, BACON can be seen as a curve-fitting program, whereas GALILEO attempts to make representational changes in order to repair faults in reasoning.

3 Proposal

Many key advances in physics have arisen from physicists' revision of their view of the world. Such change can be seen as ontology evolution, because those physicists must have updated their incorrect theories to adapt to the changing environment and goals. Human solvers can often efficiently resolve inconsistencies by revising existing theories with narrow search scopes. To formulate the reasoning patterns a creative human solver may use, a diverse range of historical records in physics will be studied. We are interested in analysing cases in which the experimental findings contradicts with some current theoretical predictions. An interesting aspect is that the theory itself is consistent, but contradicts with some new information. We provide an analysis and a formalisation of each of the case studies in an effort to help design and implement ontology repair plans.

To ensure that the repair plans to be developed are sufficiently generic and can account for a large number of historical records of ontology evolution in physics, our methodology will be a combination of

- *Development of ontology repair plans. (§3.1)*

The repair plans are compound, possibly hierarchical, systems of ontology repair operations, with associated preconditions and effects. The preconditions must be satisfied in order to execute the associated plan, and the effect defines the result of execution. These plans will be targeted at solving the theoretical/experimental contradictions and related trigger events arising in our case studies. Each plan will be generalised so that it can address a wide range of diverse situations. The composition of operations into plans will avoid the combinatorial explosions arising from a naive exhaustive search among atomic repair operations.

- *Analysis and formalisation of case studies of ontological evolution. (§3.2)*

Physics is a convenient domain for the search for case studies of ontology evolution due to the abundance of historical records of fault diagnosis and repair that lead to seminal advances. By analysing historical records, we identify patterns in the conceptual reasoning that physicists adopt to revise physical laws and theories in the light of inconsistent observations. The methodology physicists adopt to revise faulty concepts is fundamental to the design of repair plans, which essentially are composed of such patterns. Our sources will include investigations by historians, sociologists and philosophers of science. The case studies will be divided into a development and a test set.

- *Manual development of higher-order physical ontologies. (§3.2)*

As the main source of case studies is texts in, e.g., history and philosophy of science, we must translate the examples described in the texts to formal representations that are sufficiently expressive, i.e. preserving as much of the original information as possible. This may entail the development

of a new representational language. The ontologies will need to capture the historical knowledge and the higher-order physics concepts, e.g., star orbits and fluid flow. In particular, formalising the inference required to trigger ontological repair – typically deriving a contradiction between theory and experiment.

- *Development of a theory of ontology evolution. (§4.2)*

We will isolate and generalise the atomic ontology repair operations arising in our case studies. The theory provides a framework for the development of repair plans and defines properties that a repair plan should satisfy in order to be considered reasonable. For example, since repairs need to be minimal in order to avoid unmotivated and unnecessary repairs, then a suitable concept of *minimality* needs to be defined and our repairs shown to be minimal with respect to it. Much of the theory may be rationally reconstructed, based on the repair plans developed.

- *Implementation of mechanisms of inference, fault diagnosis and ontology repair. (§3.3)*

We will use a typed, polymorphic, higher-order, logic-programming language, as that seems best to represent both the physics domain and the diagnosis and repair processes. The language of implementation is higher-order, due to the many higher-order physics concepts. Moreover, many of the functions only make sense when applied to objects of certain types, e.g., the function representing gas pressure can only take a volume of gas as input.

- *Evaluation of the repaired ontologies. (§3.4)*

It is important to demonstrate the generality and the explanatory power of the repair plans, as it is undesirable to design mechanisms that are tailored to just a few similar examples. So, the repairs will be analysed with respect to the historical record of ontological evolution in our diverse range of case studies. Since we expect the size of the test set for each repair plan to be measured in tens rather than hundreds or thousands, a quantitative or statistical analysis will be inappropriate. Rather, our evaluation methodology will be based on discursive analysis.

3.1 Development of Ontology Repair Plans

Based on the analysis of the examples identified for the development set, we identify reasoning patterns that physicists use to revise theories that are inconsistent with observations. Such patterns define the function of the repair plans, the diagnosis triggers, and the repair operations. Described in this section are two example repair plans developed so far.

3.1.1 The “Where’s my stuff?” Ontology Repair Plan

Suppose we have an ontology O_t representing the current state of a physical theory and an ontology O_s representing some sensory information arising from an experiment. Suppose these two ontologies disagree over the value of some

function $stuff$ ¹ when it is applied to a vector of arguments \vec{s} of type $\vec{\tau}$. $stuff(\vec{s})$ might, for instance, be the heat content of a block of ice or the orbit of a planet.

Trigger: If $stuff(\vec{s})$ has two different values in O_t and O_s then the following formula will be triggered, identifying a potential contradiction between theory and experiment.

$$O_t \vdash stuff(\vec{s}) = v_1, \quad O_s \vdash stuff(\vec{s}) \approx v_2, \quad O_t \vdash v_1 \neq v_2 \quad (1)$$

where $O \vdash \phi$ means that formula ϕ is a theorem of ontology O . Below we deal with the case where $v_1 > v_2$. The other case is symmetric, with the roles of O_t and O_s reversed.

Split Stuff: The repair is to split $stuff$ into three new functions: visible stuff, invisible stuff and total stuff, recycling the original name for total stuff. Then we create a definition of invisible stuff in terms of total and visible stuff.

$$\forall \vec{s}:\vec{\tau}. stuff\sigma_{invis}(\vec{s}) ::= stuff(\vec{s}) - stuff\sigma_{vis}(\vec{s}) \quad (2)$$

Here, σ_{vis} and σ_{invis} are *replacements*. These resemble higher-order substitutions, except that constants, as well as variables, may be replaced with terms. When $stuff$ is a constant then σ_{vis} just replaces it with new constant standing for the visible stuff; when $stuff$ is compound the replacement is more complex, but still automatable. Similar remarks hold for σ_{invis} .

Create New Axioms: Let $\nu(O_t)$ and $\nu(O_s)$ be the repaired ontologies. We calculate the axioms of the new ontologies in terms of those of the old as follows:

$$\begin{aligned} Ax(\nu(O_t)) &::= \{ \forall \vec{s}:\vec{\tau}. stuff\sigma_{invis}(\vec{s}) ::= stuff(\vec{s}) - stuff\sigma_{vis}(\vec{s}) \} \cup \\ &\quad Ax(O_t) \\ Ax(\nu(O_s)) &::= \{ \phi \{ stuff / stuff\sigma_{vis} \} \mid \phi \in Ax(O_s) \} \end{aligned}$$

i.e. the axioms of $\nu(O_t)$ are the same as for O_t except for the addition of the new definition; the axioms of $\nu(O_s)$ are the same as for O_s except for the renaming of the original stuff to the visible stuff.

After the creation of the new axioms, the contradiction disappears. However, the theorems of the two ontologies are preserved up to renaming and the logical consequences arising from adding the new $stuff$ definition (2). Note that slight errors in the sensory data are accommodated by the use of \approx to indicate that $stuff(\vec{s})$ is *approximately* equal to v_2 in (1). If $=$ was used, then the empirical evidence must be exactly equivalent to the prediction in order to trigger the repair plan. This can be excessively strong since small empirical errors are typically acceptable in science. Note also, that $=$, \approx , $>$ and $-$ have to be polymorphic, i.e. apply to a variety of types.

Having hypothesised the existence of some hitherto invisible (i.e. not detectable by current instruments) stuff, then a natural next question is to try to develop an instrument that *can* detect it, even if indirectly.

¹ $stuff$ is a polymorphic, higher-order variable ranging over functions in physics.

3.1.2 The Inconstancy Ontology Repair Plan

Suppose we have an ontology O_t representing the current state of a physical theory and some ontologies O_s representing sensory information arising from experiments, such that different sensory ontologies give distinct values for function $stuff(\vec{s}_i)$ in different circumstances. Suppose function $V(\vec{s}_i, \vec{b}_i)$ of the i^{th} sensory ontology, where \vec{b}_i contains variables distinguishing among these circumstances, returns distinct values in each of these circumstances, but is *not* one of the parameters in \vec{s}_i , i.e. $stuff(\vec{s}_i)$ does not depend on $V(\vec{s}_i, \vec{b}_i)$. We will call $stuff(\vec{s}_i)$ the *inconstancy* and $V(\vec{s}_i, \vec{b}_i)$ the *variad*. The Inconstancy repair plan establishes a relationship between the variad $V(\vec{s}_i, \vec{b}_i)$ and the inconstancy $stuff(\vec{s}_i)$. The inconstancy might, for instance, be the gravitational constant G and the variad might be the acceleration of an orbiting star due to the gravity, which is suggested by MODified Newtonian Dynamics (MOND).

Trigger: If $stuff(\vec{s}_i)$ is measured to take different values in different circumstances, then the following trigger formulae will be matched.

$$\begin{aligned} O_s(V(\vec{s}_1, \vec{b}_1) = v_1 \dots) &\vdash stuff(\vec{s}_1) \approx c_1 \\ &\vdots \\ O_s(V(\vec{s}_n, \vec{b}_n) = v_n \dots) &\vdash stuff(\vec{s}_n) \approx c_n \end{aligned} \quad (3)$$

$$\begin{aligned} O_t &\vdash stuff(\vec{x}) ::= c(\vec{x}) \\ \exists i \neq j. O_t &\vdash stuff(\vec{s}_i) - c_i \neq stuff(\vec{s}_j) - c_j \end{aligned} \quad (4)$$

where \vec{x} can be instantiated to \vec{s}_i for $1 \leq i \leq n$, $O_s(V(\vec{s}_i, \vec{b}_i) = v_i)$ is the sensory ontology containing observations made under the condition that $V(\vec{s}_i, \vec{b}_i) = v_i$ and $V(\vec{s}_i, \vec{b}_i)$ is not an existing argument of $stuff(\vec{s}_i)$, i.e. $V(\vec{s}_i, \vec{b}_i) \notin \vec{s}_i$.

Add Variad: The repair is to change the signature of all the ontologies to relate the inconstancy, $stuff(\vec{x})$, to the variad, $V(\vec{x}, \vec{y})$:

$$\nu(stuff) ::= \lambda \vec{y}, \vec{x}. F(c(\vec{x}), V(\vec{x}, \vec{y})) \quad (5)$$

where F is a new function, whose value we will seek to determine by curve fitting against the data from the sensory ontologies.

Create New Axioms: We calculate the axioms of the new ontologies in terms of those of the old as follows:

$$\begin{aligned} Ax(\nu(O_s(V(\vec{s}_i, \vec{b}_i) = v_i \dots))) &::= \{ \phi \{ stuff / \nu(stuff)(\vec{b}_i) \} \mid \phi \in Ax(O_s(V(\vec{s}_i, \vec{b}_i) = v_i)) \} \\ Ax(\nu(O_t)) &::= \{ \phi \{ stuff / \nu(stuff)(\vec{y}) \} \mid \phi \in Ax(O_t) \setminus \{ stuff(\vec{x}) ::= c(\vec{x}) \} \} \cup \{ \nu(stuff) ::= \lambda \vec{y}, \vec{x}. F(c(\vec{x}), V(\vec{x}, \vec{y})) \} \end{aligned}$$

i.e. the axioms of $\nu(O_t)$ and the $\nu(O_s(V(\vec{s}_i, \vec{b}_i) = v_i))$ are the same as for O_t and $O_s(V(\vec{s}_i, \vec{b}_i) = v_i \dots)$ except for the replacement of the old $stuff$ with $\nu(stuff)$ and the replacement of the definition of $stuff(\vec{x})$ by the definition of $\nu(stuff(\vec{x}))$ in $\nu(O_t)$.

3.2 Analysis of Case Studies and Development of Ontologies

In this section, five case studies are presented: discovery of Latent Heat §3.2.1, discovery of dark matter §3.2.2, Modified Newtonian Dynamics §3.2.3, measurement of travel-time of light §3.2.4, and discovery of atmospheric pressure §3.2.5. For the first three cases, a formalisation of the ontologies and brief analyses of the conflicts that give rise to and the reasoning behind the evolution of ontologies are provided. A detailed informal analysis instead of a formalisation of the ontologies is presented for the last case.

3.2.1 Discovery of Latent Heat

Joseph Black discovered the concept of latent heat around 1750. Wisser and Carey [64] discuss a period when heat and temperature were conflated, which presented a conceptual barrier that Black had to overcome before he could formulate the concept of latent heat. This conflation creates a paradox: as water is frozen it is predicted to lose heat, but its heat, as measured by temperature, remains constant. Black had to split the concept of heat into energy and temperature.

The paradox faced by Black can be formalised as follows:

$$O_t \vdash \text{Heat}(H_2O, \text{Start}(\text{Freeze})) = \text{Heat}(H_2O, \text{Start}(\text{Freeze})) \quad (6)$$

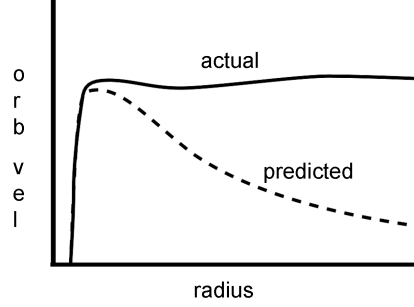
$$O_s \vdash \text{Heat}(H_2O, \text{Start}(\text{Freeze})) \approx \text{Heat}(H_2O, \text{End}(\text{Freeze})) \quad (7)$$

$$O_t \vdash \text{Heat}(H_2O, \text{Start}(\text{Freeze})) \neq \text{Heat}(H_2O, \text{End}(\text{Freeze})) \quad (8)$$

where H_2O is the water being frozen, *Freeze* is the time interval during which the freezing takes place, *Start* returns the first moment of this period and *End* the last. (6) comes from the reflexive law of equality, (7) comes from the observed constant temperature during freezing and (8) is deduced from the then current physical theory that heat decreases strictly monotonically when objects are cooled. One problem with this representation is that the proper representation of *Heat*, or specific heat, should actually depend on the *specific heat capacity* of the object, the *mass* of the object, and the *temperature difference*. That said, the two arguments in this representation, the object and the time, should be further repaired by refinement to uncover the specific dependencies. The second argument, *Time*, will still be a dependency of the three mentioned dependencies, although *Heat* itself will not depend on it. In other words, *Start(Freeze)* or *End(Freeze)* should not be an argument of the function *Heat*, but an argument of each of the functions representing the three dependencies to be uncovered. This is a promising avenue for future work.

3.2.2 Discovery of Dark Matter

The evidence for dark matter arises from various sources, for instance, from an anomaly in the orbital velocities of stars in spiral galaxies identified by Rubin [56]. Given the observed distribution of mass in these galaxies, we can use Newtonian Mechanics to predict that the orbital velocity of each star should be inversely proportional to the square root of its distance from the galactic centre (called its *radius*). However, observation of these stars show their orbital velocities to be roughly constant and independent of their radius. Figure 1 illustrates



This diagram is taken from http://en.wikipedia.org/wiki/Galaxy_rotation_problem. The x-axis is the radii of the stars and the y-axis is their orbital velocities. The dotted line represents the predicted graph and the solid line is the actual graph that is observed.

Figure 1: Predicted *vs* Observed Stellar Orbital Velocities

the predicted and actual graphs. In order to account for this discrepancy, it is hypothesised that galaxies also contain a halo of, so called, *dark matter*, which is invisible to our radiation detectors, such as telescopes, because it does not radiate, so can only be measured indirectly.

We can formalise the situation with the following formulae:

$$O_t \vdash \lambda s \in Spiral. \langle Rad(s), OrbVel(s) \rangle = Graph_p \quad (9)$$

$$O_s \vdash \lambda s \in Spiral. \langle Rad(s), OrbVel(s) \rangle \approx Graph_a \quad (10)$$

$$O_t \vdash Graph_p \neq Graph_a \quad (11)$$

where $OrbVel(s)$ is the orbital velocity of star s , $Rad(s)$ is the radius of s from the centre of its galaxy and $Spiral$ is a particular spiral galaxy, represented as the set of stars it contains. Formula (9) shows the predicted graph, $Graph_p$: the orbital velocity decreases roughly inversely with the square root of the radius (see Figure 1). This graph is deduced by Newtonian Mechanics from the observed distribution of the visible stars in the spiral galaxy. Formula (10) shows the actual observed orbital velocity graph, $Graph_a$: it is almost a constant function over most of the values of s (see Figure 1). Note the use of higher-order logic to make function objects, such as $Graph_a$, and the use of λ abstraction in (9) and (10) to create graph objects as unary functions. These two graphs are unequal (11), within the range of legitimate experimental variation.

3.2.3 Modified Newtonian Dynamics

Another explanation of the anomaly in orbital velocities of stars in spiral galaxies depicted in Figure 1 is provided by MODified Newtonian Dynamics (MOND), proposed by Moti Milgrom in 1981 as an alternative to the dark matter explanation. Essentially, MOND suggests that the gravitational constant is not a constant, but depends on the acceleration between the objects on which it is

up again. This discovery helped him come up with the theory that when Jupiter and Earth were further apart, there was more distance for light reflecting off Io to travel to Earth and therefore it took longer to reach his telescope.

Let $Time(Light)$ be the time required for light to travel. By Aristotle's theory, the time light requires to travel over any distance is zero because it is instantaneous, i.e. $Time(Light) ::= 0$, which is asserted in the original theoretical ontology, O_t . Let $O_s(Dist(p_1, p_2) = d \dots)$ be the sensory ontology describing the situation in which the distance light travels between points p_1 and p_2 is d ; in this example, D is the distance between Io and the Earth. We would like to collect observations over a reasonable range of such distances between a series of two points, P_i and P_{i+1} , where $1 \leq i \leq n$ such that $Dist(P_i, P_{i+1})$ varies within the range.

Aristotle's original theory and Roemer's theory can be represented as the following:

$$\begin{array}{ccc} O_s(Dist(P_1, P_2) = D_1 \dots) & \vdash & Time(Light) \approx T_1 \\ & \vdots & \\ O_s(Dist(P_{n-1}, P_n) = D_n \dots) & \vdash & Time(Light) \approx T_n \end{array} \quad (15)$$

$$\begin{array}{ccc} O_t & \vdash & Time(Light) ::= 0 \end{array} \quad (16)$$

$$\begin{array}{ccc} \exists i \neq j. O_t & \vdash & Time(Light) - T_i \neq Time(Light) - T_j \end{array} \quad (17)$$

where O_t and O_s represent Aristotle's theory and Roemer's observations respectively; $Dist(P_i, P_j)$ returns the straight line distance between points P_i and P_j ; $Time(Light)$ gives the travel time of light, treating light as an object. (15) comes from Roemer's observations of the different travel of times of light in different occasions, such that in each of these occasions the straight line distance between the Earth and Io were different, ranging from D_1 to D_n . (16) comes from Aristotle's original theory that light travels instantaneously, thus light has a travel time of 0s. (17) shows a variation in the travel times observed.

3.2.5 Discovery of Atmospheric Pressure

In Aristotelian physics, a vacuum was believed to be impossible to produce. It was believed that pumping of water was a consequence of water immediately replacing the removed matter, because there could be no vacuum. Galileo challenged this idea and insisted that vacuum is in fact possible to produce, but concluded wrongly that the pumping of water was a consequence of a "force" or a "resistance" exerted by the vacuum *alone*. The experiment performed to support this explanation was to uncompress an water-filled apparatus, similar to a syringe with the needle end blocked. By Galileo's theory, the uncompression is resisted by the vacuum alone, i.e. that only the amount of stuff in a vacuum inside the barometer *causes* the mercury level to vary. Evangelista Torricelli, a follower of Galileo and generally credited for the invention of the barometer, noticed the level of the fluid in his experimental setup changed slightly each day. Torricelli's observations contradicted Galileo's theory because the level of the fluid changed even when the vacuum space, which Galileo believed to be the sole cause, was not disturbed. Torricelli's key to the correct explanation is

the discovery of a changing atmospheric pressure and its effect on the fluid, of which Galileo was not aware.

Galileo’s explanation is not wrong but is a partial explanation, because he assumed a constant atmospheric pressure. Torricelli’s explanation is clearly more acceptable because it is more general, thus more informative. These are examples of *contextual causal statements*, in which causes and effects can be seen as *differences* to some background conditions, or the normality [34]. If the background conditions are different, the causal explanations may be different. In perspective, each of Galileo’s and Torricelli’s interpretations of the problem has different background conditions: Galileo assumed a constant atmospheric pressure, whereas Torricelli did not. As a result, the background conditions influenced the causal explanations in such a way that Torricelli concluded the cause to be both the vacuum and the atmospheric pressure, whereas Galileo concluded it to be the vacuum alone as he believed the atmospheric pressure cannot change.

Based on the idea of contrasting a restricted set of events bounded according to the context, Cheng and Novick introduce the notion of *focal sets* in [9]. The focal set for a contrast is the contextually determined set of events that a reasoner selects as being relevant to the given causal event. The focal set is typically a subset of the *universal set*, which contains all the events of which the reasoner is aware. Cheng and Novick define a *cause* to be a factor that noticeably increases the likelihood of the effect. Clearly, the theory of focal sets assumes probabilistic causation [47]. At this stage, however, our particular interest is not on the details of probabilistic causation, so we only look for covariation between a factor and the effect in question based on a particular focal set. Fortunately for this example, even if probabilities are not explicitly taken into consideration, little faithfulness to the underlying history and accuracy of the focal sets of Galileo and Torricelli are lost.

Galileo’s and Torcelli’s focal sets are depicted in Figure 2, where the events highlighted in bold represent those producing the effect that the level of the mercury holds at some marked level h at a time moment T , i.e. $Height(Hg, T) = h$. As Galileo was not aware of the possibility of having different atmospheric pressures, his focal set contains only those events where the atmospheric pressure is fixed, i.e. $AtmospP(T) = p$. Given a fixed volume of fluid, he believed the only way for the fluid to reach height h is to let the amount of vacuum be v , i.e. $Height(Hg, T) = h$ only if $Vacuum(T) = v$. Based on this setup, $Vacuum(T) = v$ should be perceived as a cause of the effect because it covaries with the effect within Galileo’s focal set. As for $AtmospP(T)$, it should *not* be perceived as a cause given the focal set because its values are constant throughout the set, thus it does not covary with the effect. On the contrary, Torricelli believed that the atmospheric pressure can vary and observed a variation in the level of the fluid without altering the amount of vacuum. So, his focal set contains the events where the level of the fluid changes when the atmospheric pressure is not p and the amount of vacuum is maintained, i.e. $Height(Hg, T) \neq h$ when $AtmospP(T) \neq p$ and $Vacuum(T) = v$. Galileo’s focal set is a subset of Torricelli’s because Torricelli agreed with Galileo that the amount of vacuum *can* cause a change in the fluid level. For Torricelli’s focal set, the values of $AtmospP(T)$ are not constant, so both $AtmospP(T) = p$ and $Vacuum(T) = v$ are jointly sufficient for the occurrence of the effect.

The formalisation of this example is still work to be done. The key elements

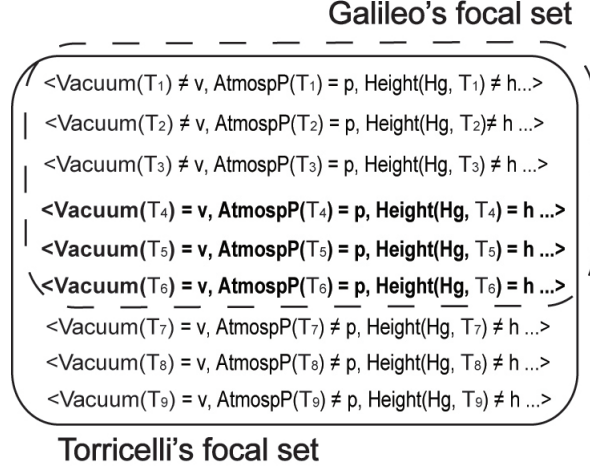


Figure 2: Galileo's (dashed) and Torricelli's (solid) focal sets for explaining the “force” exerted on the mercury in a barometer, with events producing the effect $Height(Hg) = h$ highlighted in bold.

to be represented are the discrepancies in the background knowledge that led Galileo and Torricelli to different conclusions. So, the idea of focal sets will be further investigated in order to illustrate the influence of contextual knowledge on causal statements.

3.3 Implementation

Many physical concepts are essentially higher-order, for instance, the orbit of a star, which can be represented as a function object returning the 3-dimensional position in space for a given time moment. Of course, it can be argued that an orbit can be treated as a first-order object, represented as the set of all coordinates for all time moments. Such representation would be unnecessarily complex and un-descriptive. Moreover, many of the functions only make sense when applied to objects of certain types, e.g., the function representing air pressure can only be applied to a volume of gas.

Given the higher-order requirements, we have chosen λ Prolog [42] as our implementation language because it provides a polymorphic, higher-order logic. This is well suited to the representation of the higher-order functions that occur in both the object-level domain of physics and the meta-level formulae of the repair plans. It provides higher-order unification to support matching of the meta-level triggers of the repair plans to the object-level triggering formulae in the case studies. It facilitates the representation of polymorphic functions such as $-$, $<$, \approx and $=$. These features combined to support rapid prototyping of the repair plans and their application to the development case studies.

To illustrate the implementation, here is the main clause³ of the WMS plan.

```
repair O1 O2 NA1 NA2 :-
```

³Confusingly, λ Prolog uses the convention that words representing variables start with upper-case letters and constants with lower-case, which is the inverse of the standard mathematical convention we have used for our mathematical formula.

```

% Repair triggered. Find stuff, args and parity
wms_trigger 01 02 S L P,
% Pick replaced stuff from S or L
choose S L Tot,
% Calculate total, visible and invisible stuff
newstuff S L Tot STot SVis SInvis,
% Get original axioms
axioms 01 A1,
% of both ontologies
axioms 02 A2,
% Flip to find opposite parity
flip P FP,
% Change both sets of axioms
change P 01 A1 STot SVis SInvis NA1,
change FP 02 A2 STot SVis SInvis NA2.

```

01 and 02 are the input initial theoretical and experimental ontologies (but which is which depends on the parity P). NA1 and NA2 are their output repaired axioms. `wms_trigger` checks that the triggering formula (1) is matched and returns the instantiations of *stuff*, its list of arguments and a parity according to whether $v_1 > v_2$ or $v_1 < v_2$. `choose` picks a candidate *Tot* to be replaced in σ_{vis} and σ_{invis} , and `newstuff` uses these replacements to calculate the new total, visible and invisible *stuff*. The old axioms are then found by `axioms` and repaired into the new axioms by `change`.

We are currently experimenting with Terzo and Teyjus implementations of λ Prolog. We have experimented with Twelf, but unfortunately it had insufficient support for polymorphism, which is vital to our work. We continue to investigate other potential platforms to identify one that both provides the functionality we require and is well maintained. A possible alternative to implementing in λ Prolog is to use Isabelle [44], which is a generic framework for proving higher-order theorems. However, Isabelle uses a procedural proof style, which may make coding the repair plans more difficult.

3.4 Evaluation of Repaired Ontologies

We will be looking specifically for generality and explanatory power from our repair plans. To this end, we will seek diversity in our test set and emergent abstraction from the uniform processing of apparently diverse case studies. Each example will be manually identified, represented and encoded, so the test set will likely contain only multiples of tens of examples. To demonstrate the proposed method for evaluation of repaired ontologies, the four examples described in §3.2.1, §3.2.2, §3.2.3 and §3.2.4 are to be applied to the WMS and Inconstancy repair plans in this section.

3.4.1 Application of WMS to the Discovery of Latent Heat

The formulae (6), (7) and (8) match the WMS repair plan trigger (1) in §3.1.1 with the following substitution:

$$\{Heat/stuff, \langle H_2O, Start(Freeze) \rangle / \vec{s}, Heat(H_2O, Start(Freeze)) / v_1, Heat(H_2O, End(Freeze)) / v_2\}$$

To effect the repair we will define $\sigma_{vis} = \{Temp/stuff\}$ and $\sigma_{invis} = \{Lhf/stuff\}$, respectively, in anticipation of their intended meanings, where *Lhf* can be read as the latent heat of fusion. These choices instantiate (2) in Appendix 3.1.1 to:

$$\forall o:obj, t: mom. Lhf(o, t) ::= Heat(o, t) - Temp(o, t)$$

which is not quite what is required, but is along the right lines. Some further indirect observations of *Lhf* are required to witness its behaviour under different states of *o* so that it can be further repaired, e.g., the removal of its *t* argument. The *Temp* part of the new definition needs to be further refined so that its contribution of energy depends both on temperature and mass. These further refinements will be the subject of future ontology repair plans.

In the repaired ontologies, since $Heat(H_2O, Start(Freeze))$ is greater than $Heat(H_2O, End(Freeze))$, the repaired triggering formulae are transformed to :

$$\begin{aligned} \nu(O_t) &\vdash Heat(H_2O, Start(Freeze)) = Heat(H_2O, Start(Freeze)) \\ \nu(O_s) &\vdash Temp(H_2O, Start(Freeze)) = Temp(H_2O, End(Freeze)) \end{aligned}$$

which breaks the derivation of the detected contradiction, as required.

3.4.2 Application of WMS to Dark Matter

The formulae (9), (10), and (11) instantiate the WMstrigger preconditions (1) with the following substitution:

$$\{\lambda s \in g. \langle Rad(s), Orb.Vel(s) \rangle / stuff, \langle Spiral \rangle / \vec{s}, Graph_p / v_1, Graph_a / v_2\}$$

Note that the repair plan works perfectly well with higher-order objects as the values v_1 and v_2 , provided that polymorphic $-$ and \neq can be defined as having meaning over this data-type: in this case a piecewise subtraction over the individual values for each star and a fuzzy, negated equality (\approx) between graphs.

To effect the repair we will define $\sigma_{vis} = \{Spiral_{vis}/g\}$ and $\sigma_{invis} = \{Spiral_{invis}/g\}$, so the instantiation of definition (2) suggested by this triggering is:

$$\begin{aligned} \lambda s \in Spiral_{invis}. \langle Rad(s), Orb.Vel(s) \rangle \\ ::= \lambda s \in Spiral. \langle Rad(s), Orb.Vel(s) \rangle - \\ \lambda s \in Spiral_{vis}. \langle Rad(s), Orb.Vel(s) \rangle \end{aligned}$$

where $Spiral_{vis}$ is the visible part of the galaxy, that can be detected from its radiation, and $Spiral_{invis}$ is its dark matter part.

In the repaired ontologies, since $Graph_p < Graph_a$, the repaired triggering formulae are:

$$\begin{aligned} \nu(O_t) &\vdash \lambda s \in Spiral_{vis}. \langle Rad(s), Orb.Vel(s) \rangle = Graph_p \\ \nu(O_s) &\vdash \lambda s \in Spiral. \langle Rad(s), Orb.Vel(s) \rangle \approx Graph_a \end{aligned}$$

which breaks the previous derivation of a contradiction, as required. Note that, unlike the latent heat case study, it is the repaired *theoretical* ontology that formalises the visible stuff and it is the repaired *experimental* ontology that formalises the total stuff. This is because $Graph_p$ is based on the visible mass in the spiral and is, therefore, smaller than $Graph_a$.

3.4.3 Application of Inconstancy to Modified Newtonian Dynamics

The formulae (12), (13) and (14) match the Inconstancy repair plan trigger (3), (4), and (5) in §3.1.2 with the following substitution:

$$\begin{aligned}
O_s(Acc(S_1) = \\
A_1 \dots) \quad &\vdash \quad G = M2OV^{-1}(OV(S_1), Mass(S_1), \\
&\quad \lambda s \in Spiral \setminus \{S_1\}. \langle Posn(s), Mass(s) \rangle) \\
&\quad (= G_1) \\
&\quad \vdots \quad \quad \vdots \\
O_s(Acc(S_n) = \\
A_n \dots) \quad &\vdash \quad G = M2OV^{-1}(OV(S_n), Mass(S_n), \\
&\quad \lambda s \in Spiral \setminus \{S_n\}. \langle Posn(s), Mass(s) \rangle) \\
&\quad (= G_n) \\
O_t \quad &\vdash \quad G ::= 6.67 \times 10^{-11} \\
\exists i \neq j. O_t \quad &\vdash \quad G - G_i \neq G - G_j
\end{aligned}$$

The formulae above triggers the plan with the following substitution:

$$\begin{aligned}
\{G/stuff, \langle \rangle/\vec{s}_i, \langle \rangle/\vec{s}_j, \langle \rangle/\vec{x}, 6.67 \times 10^{-11}/c, \\
Acc/V, \langle S_i \rangle/\vec{b}_i, G_1/c_1, G_n/c_n\}
\end{aligned}$$

Since G is a constant, \vec{s}_i , \vec{s}_j and \vec{x} are simply empty vectors.

Following the instructions for repair, the variad is given to the inconstancy by:

$$\nu(G) ::= \lambda s. F(6.67 \times 10^{-11}, Acc(s))$$

and the repaired triggering formulae are therefore:

$$\begin{aligned}
\nu(O_s(Acc(\\
S_1) = A_1 \dots)) \quad &\vdash \quad \nu(G)(S_1) = M2OV^{-1}(OV(S_1), \\
&\quad Mass(S_1), \lambda s \in Spiral \setminus \{S_1\}. \\
&\quad \langle Posn(s), Mass(s) \rangle) (= G_1)
\end{aligned} \tag{18}$$

$$\begin{aligned}
&\quad \vdots \quad \quad \vdots \\
\nu(O_s(Acc(\\
S_n) = A_n \dots)) \quad &\vdash \quad \nu(G)(S_n) = M2OV^{-1}(OV(S_n), \\
&\quad Mass(S_n), \lambda s \in Spiral \setminus \{S_n\}. \\
&\quad \langle Posn(s), Mass(s) \rangle) (= G_n) \\
\nu(O_t) \quad &\vdash \quad \nu(G) ::= \lambda s. F(6.67 \times 10^{-11}, Acc(s))
\end{aligned} \tag{19}$$

which breaks the derivation of the detected contradiction, as required.

The function F can be determined by finding the best-fit curve for the whole dataset, in which each data point represents an observed G_i made under a particular condition $Acc(S_i) = A_i$. F is a reasonable approximation only if a fairly large number of observations of G_i for a wide range of accelerations of stars $Acc(S_i)$ are analysed. If F is a correct and complete approximation of $\nu(G)$,

then $F(6.67 \times 10^{-11}, Acc(s))$ returns the unrepaired value 6.67×10^{-11} if a star s has an acceleration much greater than $1.2 \times 10^{-10} ms^{-2}$ (close to the centre of the galaxy). If s has an acceleration that is much less than $1.2 \times 10^{-10} ms^{-2}$ (near the periphery of the galaxy), the value returned will be greater than 6.67×10^{-11} and proportional to $Acc(s)^2 \times Rad(s)^2$, where $Rad(s)$ is the radius of the star's orbit.

Clearly, the repair performed to give (18) and (19) is very different from that performed to give the dark matter theory – respectively, the acceleration of a star becomes a dependent variable of the repaired definition of the Gravitational Constant and the applicability of the repaired Newtonian theory is limited to only the observable part of the galaxy.

3.4.4 Application of Inconstancy to the Travel Time of Light

The discrepancy that arise between Aristotle's theory and Roemer's observations (§3.2.4) is also another example of Inconstancy, as Roemer's observed times unexpectedly varied when the Aristotle's theory expected them to remain constant. The formulae (15), (16), and (17) instantiate the trigger formulae with the following substitution:

$$\{Time(Light)/stuff, \langle \rangle / \vec{s}_i, \langle \rangle / \vec{s}_j, \langle \rangle / \vec{x}, 0/c, \\ Dist/V, \langle P_i, P_{i+1} \rangle / \vec{b}_i, T_1/c_1, T_n/c_n\}$$

To insert the variad, $Dist(Light, T_i)$, into the new definition, we follow (5) to get:

$$\nu(Time(Light)) ::= \lambda p_1, p_2. F(0, Dist(p_1, p_2))$$

The repaired ontologies are therefore:

$$\begin{aligned} & \nu(O_s(Dist(P_1, \\ & P_2) = D_1 \dots)) \vdash (\nu(Time(Light))(P_1))(P_2) = D_1 \\ & \quad \vdots \quad \quad \quad \vdots \\ & \nu(O_s(Dist(P_{n-1}, \\ & P_n) = D_n \dots)) \vdash (\nu(Time(Light))(P_{n-1}))(P_n) \\ & \quad \quad \quad = D_n \\ & \nu(O_t) \vdash \nu(Time(Light)) ::= \lambda p_1, p_2. F(0, Dist(p_1, p_2)) \end{aligned}$$

which together resolve the detected contradiction. The new definition of $Time(Light)$, $\nu(Time(Light))$, is what is required by the general definition of time of travel, which can be put simply that if an object, including a stream of photons, travels at a finite speed, then the time of travel depends on the distance, i.e. $Time \propto Distance$.

3.5 Contribution to Knowledge

When agents communicate or exchange information between each other, the prerequisite is that a consensus on the ontological concepts involved has to be reached between them. Unfortunately, agents in a vast distributed environment, such as the Semantic Web, tend to capture different parts of the world, or the

same part differently. This results in inconsistent representations in their ontologies, partly caused by the lack of a controlled or centralised design scheme. Such a design scheme is commonly conceded to be extremely difficult, or at least highly undesirable, to enforce in a large, heterogeneous environment. It is therefore extremely useful to enable agents to autonomously repair their own ontologies in the light of external conflicting information. Proper and reliable communication can be achieved only if agents themselves are capable of updating their own ontologies in a sound fashion. Humans can detect faults in theories and revise faulty theories with ease when necessary. In order for machines to perform a similar task, our approach is to design ontology repair plans that imitate a human’s ability to revise theories and repair conceptual faults. Repair plans aim at detecting inconsistencies between an agent’s current ontology and the newly obtained information. If inconsistency is detected, repair plans guide the manipulation process so that the inconsistency is resolved. We are interested in designing generic repair mechanisms, which are not tailored for any particular ontological representation language.

A formal theory of ontology evolution will define desirable properties that a reasonable repair plan should satisfy. Such a theory will prevent us from tackling the problem in an *ad-hoc* manner. The theory serves a similar purpose to that of the AGM postulates, which define what constitutes rational belief revision. The formal theory could help the overall development in ontology evolution research as the AGM framework has in belief revision. Moreover, the patterns that compose the diagnosis and repair operations are based on historical cases of ontology evolution in physics. So, the identification of such patterns requires detailed analyses of the theory revision skills physicists have applied throughout history. This compilation of ontology evolutions in physics will illuminate our understanding of theory revision and aid future research in automated conceptual reasoning, and in other disciplines, e.g., cognitive science and philosophy of science.

4 Future Direction

4.1 Overview

We envision a collection of additional repair plans that focus on resolving other faults in conceptual reasoning. For instance, scientists often apply causal reasoning when defining processes as mathematical functions. Physics formulas do not explicitly describe causal relationships among variables, because each variable only represents a sufficient condition. Some philosophers, e.g., Russell [57], have argued that fundamental sciences, e.g., gravitational astronomy, do not talk of causes, but causal notions are widely believed to still be a best way to help physicists conceive the world. Norton [45] agrees with Russell and argues that there is no central principle of causality in fundamental sciences, but he urges that qualitative causal conceptions provide a convenient way to understand otherwise opaque processes. An example ontology evolution is the discovery of atmospheric pressure, in which Galileo and Torricelli set the causal question in different contexts by making different assumptions about the atmosphere (§3.2.5). Consequently, the different contexts resulted in inconsistent causal explanations.

Another interesting focus is to design a repair plan to handle analogical reasoning to help explain new scientific discoveries and derive new theories [62]. Essentially, an analogical consideration typically involves determining whether two seemingly unrelated concepts are isomorphic. The main challenge here is to discover interesting properties and determine their interestingness, and to formalise the (pseudo-) isomorphism. In mathematics, an example is the translation of theorems in geometry into Cartesian algebra. An instance from the physics domain could be Galileo's and, again, Roemer's measurement of the travel time of light. Roemer's measurements can be seen as a successful attempt of Galileo's initial experiment, which involved two shutter lanterns across two hilltops. Roemer's attempt is a close analogical version of Galileo's since both involved measuring light's travel time between two points, except Roemer used Jupiter as the shutter and Io as the light source over a far greater distance.

Just as the AGM framework is important to the research in belief revision operators, a formal theory of ontology evolution is vital to the progress of development of ontology evolution mechanisms. Such a theory will define properties that an operation of ontology repair should satisfy in order to be sound, similar to how the AGM postulates describe those that a belief revision operators should satisfy in order to be rational. The postulates described in §4.2 appear to be incompatible with ontology evolution, but the underlying spirit can become the basis of parts of the theory of ontology evolution. Furthermore, the foundation of the theory can be built on discussions in philosophy of science and investigations in scientific reasoning. The fruitful literature in belief revision can also be helpful, as other issues that arise in belief revision might be closely connected to those potentially faced by ontology evolution. Philosophy of belief revision could therefore be helpful for us to identify potential issues and the corresponding challenges to overcome.

4.2 Development of a Theory for Ontology Evolution

Ontology evolution presents exciting theoretical challenges. For instance, to define a concept of minimal repair we are experimenting with extending *conservative extension* to changes of signature, namely:

$$\phi \in \mathcal{L}(O) \implies (\nu(O) \vdash \nu(\phi) \iff O \vdash \phi) \quad (20)$$

where $\phi \in \mathcal{L}(O)$ means that ϕ is a formula in the language of ontology O and $\nu(\phi)$ is the repaired ϕ in repaired ontology $\nu(O)$. In the repair plan in §3.1.1 both $\nu(O_t)$ and $\nu(O_s)$ are conservative in this extended sense. Their combination, of course, is not, since the purpose of the repair is to prevent a contradiction being derived in the repaired combination.

The theory for ontology evolution may draw ideas from that for belief revision, e.g., the AGM framework. Of course, the proposed approach of ontology repair is not an example of contraction, i.e. the removal of a belief from an ontology. Thus, the AGM contraction postulates obviously cannot be adopted wholesale. Even for AGM revision, the fundamentals between belief revision and ontology evolution can be very different. In the rest of this section, we provide a crude attempt at a reformulation of the KM postulates for the ontology context.

The KM postulates, outlined in §2.6, are based on the representation $K = \{\phi \mid \psi \vdash \phi\}$, where ψ is a propositional formula representing a knowledge-base

and ϕ is a sentence. In the context of ontology repairs, we can simply treat the ontology O_t as ψ and the sensory ontology O_s as μ , because repairing O_t in the light of O_s takes into account the new information provided by O_s in a somewhat similar way to belief revision. For a more intuitive comparison, we use a simpler syntax: Instead of using ν , here \bullet is used to represent an ontology merge operator merging two ontologies together, possibly partially. The outputs of ν and \bullet are intended to be the same, i.e. a repaired ontology which has accommodated some new information. The function of the ontology merge operator \bullet can be seen as being similar to that of a belief operator \circ , which merges an existing knowledge-base with some newly acquired information resulting in a change of beliefs. Below are the KM postulates reformulated in terms of ontologies O_t and O_s and the ontology merge operator \bullet below; this is an attempt to share much of the spirits of KM:

(KM-O1) $O_s \vdash \phi \implies O_t \bullet O_s \vdash \phi$. *All new sensory knowledge is retained.*

(KM-O2) $O_t \cup O_s \not\vdash \perp \implies (O_t \bullet O_s \equiv O_t \cup O_s)$ *Merge the ontologies naively whenever such an ontology is consistent.*

(KM-O3) $O_t \not\vdash \perp \wedge O_s \not\vdash \perp \implies O_t \bullet O_s \not\vdash \perp$. *The repair plan should resolve any inconsistency arising in the merge and not introduce unwarranted inconsistency.*

(KM-O4) $[O_{t_1} \equiv O_{t_2} \wedge O_{s_1} \equiv O_{s_2}] \implies [O_{t_1} \bullet O_{s_1} \equiv O_{t_2} \bullet O_{s_2}]$. *The repair plan should abide by the principle of Irrelevance of Syntax.*

Since KM1 is equivalent to R2 and $K * \mu = \{\phi \mid \nu(O_t) \vdash \phi\}$, $\mu \in K * \mu$ if and only if every theorem of O_s is in the theory of $\nu(O_t)$. However, KM-O1 is not desirable because the aim of repair plans is to *repair* inconsistent ontologies so that inconsistencies can be resolved and not necessarily to *learn* from sensory information. Further, experimental evidence is often questionable, so it may not be desirable to retain everything observed. Overall, KM-O1 enforces that the knowledge represented in the current ontology should be at least as updated as the sensed. With this in mind, KM-O1 can be revised:

(KM-O1') $\phi \in \mathcal{L}(O) \implies [\nu(O) \vdash \nu(\phi) \iff O \vdash \phi]$

Clearly, KM-O1' exploits the definition of conservative extension (20). KM-O1' shares a similar spirit to that behind KM-O1 and KM1, i.e. to govern the sort of theorems the repaired ontologies/knowledge-base could contain. It enforces that $\nu(O)$ must be a conservative extension of O . However, confining all repaired ontologies to those that are conservative extensions of their originals can often be excessively strong. We believe it should be reasonable to repair ontologies without extending the original signature, and so more work is required to revise KM-O1'.

KM-O2 follows the intuition that the obvious path is taken if there is no conflict. In belief revision, such a path is to take the conjunction of the existing belief and the new. In ontology repair, it would be unnecessary to add and learn all new concepts. Therefore, a reasonable repair plan should allow an ontology to retain all concepts that are *not* false:

(KM-O2') $O \not\vdash \perp \implies \nu(O) \equiv O$

Clearly, KM-O2' is adapted from KM-O2, which is a direct translation of KM2. KM-O2' tries to share the same underlying spirit of KM-O2, i.e. that the trivial path is taken if there is no conflict. With the repair plans so far developed (§3.1), it may be safe to assume to impose no change to consistent ontologies, i.e. the repaired version of a consistent ontology is equivalent to itself. However, we may want to design repair plans to not eliminate inconsistency, but to, e.g., repair unexpected regularities and merging/splitting representations. It may appear that there may not be a trivial path since the repair plans are envisioned to resolve a wide variety of faults, ranging from those arising from logical inconsistency to inefficiency in inference. If we give a classification of repair plans, then a trivial path for each class may become more appropriate and obvious.

As in belief revision operators, the key problem in repair plans is maintaining consistency; both belief revision operators and repair plans should not introduce any unwarranted inconsistency. However, the criteria for maintaining consistency are different for belief revision and ontology evolution; therefore, KM-O3 would need to be radically revised. For example, a repaired ontology is not always guaranteed to be consistent because only one type of inconsistency, the type for which the repair plan is designed, would have been eliminated.

Finally, KM-O4 can be represented in terms of ν as:

$$(KM-O4') [O_i \equiv O_j] \implies [\nu(O_i) \equiv \nu(O_j)]$$

However, ontology evolution considers signature evolution as well, so that the representation can change. Unlike in belief revision, syntactical form is *not* irrelevant in ontology evolution.

Although the AGM postulates (and the reformulations) do not seem to fit the repair operations we want, these along with other formalisations of related issues in belief change are worth further exploration.

5 Programme of Work

Our discussions of previous work (§2), central ideas (§3) and future directions (§4) have provided much of the detail of our programme of work, so they will not be repeated. Instead, we give only an outline of our workplan by dividing it into a series of iterative, interdependent workpackages, which will be carried out in parallel. The Gantt Chart is shown in Figure 3.

WP 1. Analysis and Formalisation of Examples of Ontology Evolution

The objective of WP 1 is to discover examples from the history of physics and to formalise them in higher-order logic. More specifically, this workpackage will focus on identifying and formalising candidate case studies from the history of physics. More details are described in §3.2.

Duration: 4 months spread over 24 months. **Deliverables:** Analysis on and formalisations of examples of contradictory prior ontologies. **Dependencies:** None.

WP 2. Design of Repair Plans Based on the analysis and the formalisations of the development set of examples of ontology evolution to be delivered in

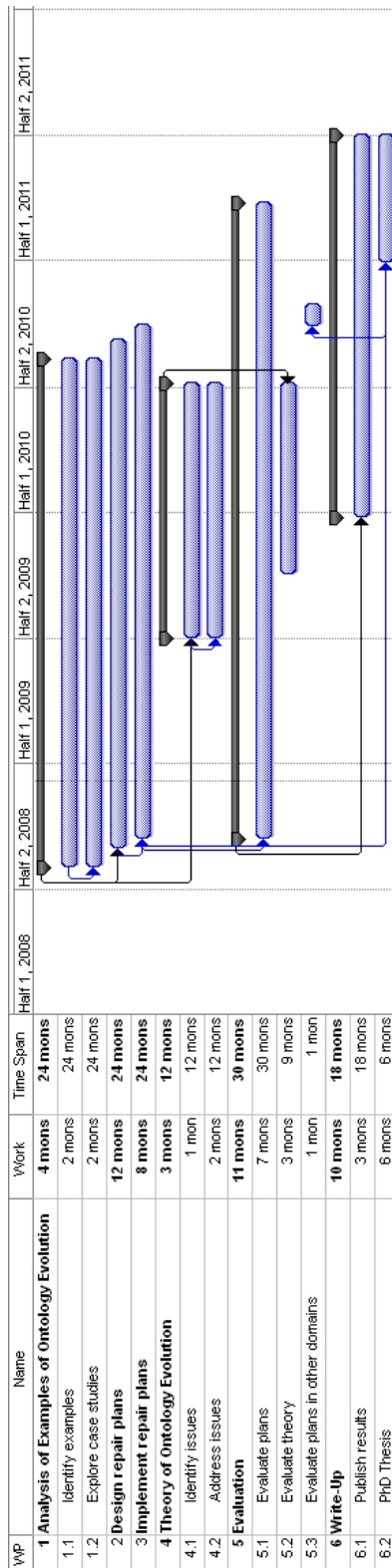


Figure 3: Gantt Chart describing all WP 1 to WP 6.

WP 1, we will design up to 10 generic repair plans. More details are described in §3.1.

Duration: 12 months spread over 24 months. **Deliverables:** Up to 10 repair plans. **Dependencies:** Formulation of development set in WP 1 and analysis of the ontology evolution illustrated by these examples.

WP 3. Implementation of Repair Plans The various repair plans and example ontologies will be implemented in a higher-order, logic-programming language. The repair plans will be tested on the development set delivered in WP 1 until they are performing dependably to specification. More details are described in §3.3.

Duration: 8 months spread over 24 months. **Deliverables:** An implemented system of repair plans and example ontologies. **Dependencies:** WPs 1, 2 and 4.

WP 4. Theory of Ontology Evolution This workpackage will involve the development of a theory addressing the fundamental issues of ontology evolution, including rationality and consistency maintenance.

Duration: 3 months spread over 12 months. **Deliverables:** A theory of ontology evolution. **Dependencies:** WPs 1, 2 and 3.

WP 5. Evaluation of Plans The aim of this workpackage is to evaluate the repair plans by their generality and explanatory power, to examine the correctness of our theory and to explore new application domains.

Duration: 11 months spread over 30 months. **Deliverables:** Analysis and formalisation of example repaired ontologies. **Dependencies:** WPs 1, 2, 3 and 4.

WP 6. Write-Up The publication of the results of the research will be spread over the second half of the whole period, while a PhD thesis will be the focus over the final 6 months.

Duration: 10 months spread over 18 months. **Deliverables:** PhD Thesis, various conference and journal papers. **Dependencies:** All other workpackages.

5.1 Potential Risks

Because this project is novel, ambitious and adventurous basic research, it is inevitable that it will not proceed exactly to plan. We have, therefore, not proposed the kind of detailed, tightly-organised workplan that is required in a development project. Rather, we have proposed ambitious goals within a broad-brush workplan, and expect to dynamically replan this as our research reveals more clearly the difficulties and opportunities we face.

The major risk faced by the project is that nearly all ontology evolution in physics is found to be too diverse to be emulated by a small set of generic and widely applicable repair plans. Our initial investigations are promising, but if this risk materialises then we will focus our efforts on the areas of the domain on which our techniques are most effective and will try to characterise both those successful areas and those on which our techniques are unsuccessful

with the intention of gaining a deeper understanding of the forces at work. We will also use our theory of ontology evolution to apply sequences of atomic repair operations in situations where repair *plans* are unobtainable. We will seek alternative heuristics to control the resulting combinatorial explosion.

If we cannot identify an ideal implementation language, we will err on the side of a well-maintained language and augment it with any missing functionality ourselves.

6 Conclusion

Further progress in handling ontology evolution is now urgent, due to the demand created by multi-agent systems. We argue that agents must be able to update their own ontologies to resolve inconsistencies in order for proper communication in a distributed, heterogeneous environment. By analysing the abundant historical records of ontology evolution in physics, we design diagnosis and repair operations based on the method adopted by physicists for revising theories and resolving experimental contradictions. The atomic operations are then composed together to achieve greater explanatory power. A formal theory of ontology evolution will connect the repair plans together by defining desirable properties that should hold, e.g., soundness. This project, therefore, represents an important step toward equipping agents with automated ontology repair mechanisms and formalising the criteria of correct, reasonable repair operations.

References

- [1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50(2):510–530, June 1985.
- [2] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, et al. Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25:25–29, 2000.
- [3] F. Baader, D. Calvanese, D.L. McGuinness, D. Nardi, and P.F. Patel-Schneider. The description logic handbook: theory, implementation, and applications. 2003.
- [4] O. Bodenreider. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32(90001):267–270.
- [5] E. Bozsak, M. Ehrig, S. Handschuh, A. Hotho, A. Maedche, B. Motik, D. Oberle, C. Schmitz, S. Staab, L. Stojanovic, et al. KAON-Towards a Large Scale Semantic Web. In *Proceedings of the Third International Conference on E-Commerce and Web Technologies*, pages 304–313. Springer-Verlag London, UK, 2002.
- [6] D. Brickley and R.V. Guha. RDF Vocabulary Description Language 1.0: RDF Schema. *W3C Recommendation*, 2004.

- [7] A. Bundy and F. McNeill. Representation as a fluent: An AI challenge for the next half century. *IEEE Intelligent Systems*, 21(3):85–87, 2006.
- [8] C. Castelfranchi. Guarantees for autonomy in cognitive agent architecture. *Proceedings of the workshop on agent theories, architectures, and languages on Intelligent agents table of contents*, pages 56–70, 1995.
- [9] P. W. Cheng and L. R. Novick. Causes versus enabling conditions. *Cognition*, 40(1-2):83–120, August 1991.
- [10] A. Darwiche and J. Pearl. On the logic of iterated belief revision. *Artificial Intelligence*, 89(1-2):1–29, 1997.
- [11] A. Doan, J. Madhavan, P. Domingos, and A. Halevy. Ontology matching: A machine learning approach. *Handbook on Ontologies in Information Systems*, pages 397–416, 2004.
- [12] George F. R. Ellis and Jean-Philippe Uzan. ‘c’ is the speed of light, isn’t it? *AM.J.PHYS.*, 73:240, 2005.
- [13] G. Flouris, D. Manakanatas, H. Kondylakis, D. Plexousakis, and G. Antoniou. Ontology change: classification and survey. *The Knowledge Engineering Review*, 23(02):117–152, 2008.
- [14] M. Freund and D. Lehmann. Belief Revision and Rational Inference. *Technical Report TR94-16, Hebrew University*, 2002.
- [15] N. Friedman and J.Y. Halpern. Belief Revision: A Critique. *Journal of Logic, Language and Information*, 8(4):401–420, 1999.
- [16] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider. Sweetening Ontologies with DOLCE. *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, pages 166–181, 2002.
- [17] P. Gärdenfors and D. Makinson. Revisions of knowledge systems using epistemic entrenchment. In *Proceedings of the 2nd conference on Theoretical aspects of reasoning about knowledge*, pages 83–95. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 1988.
- [18] P. Gärdenfors and H. Rott. Belief revision. In D. Gabbay, C. J. Hogger, and J.A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 4, pages 35–132. Oxford Science Publications, 1995.
- [19] M. R. Genesereth and N. J. Nilsson. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann, Palo Alto, CA., 1987.
- [20] F. Giunchiglia and P. Shvaiko. Semantic matching. *The Knowledge Engineering Review*, 18(03):265–280, 2004.
- [21] Thomas R. Gruber. A translation approach to portable ontology specification. *Knowledge Acquisition*, 5(2), 1993.

- [22] P. Haase, F. van Harmelen, Z. Huang, H. Stuckenschmidt, and Y. Sure. A Framework for Handling Inconsistency in Changing Ontologies. In *The Semantic Web-ISWC 2005: 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005: Proceedings*. Springer Verlag, 2005.
- [23] P. J. Hayes. The naïve physics manifesto. In *Expert Systems in the Micro-electronic Age*, pages 242–70. Edinburgh University Press, 1979.
- [24] J. Hendler. Agents and the Semantic Web. *IEEE Intelligent Systems*, 16(2):30–37, 2001.
- [25] R. Hoekstra, J. Breuker, M. Di Bello, and A. Boer. The LKIF Core Ontology of Basic Legal Concepts. In *Proceedings of LOAIT07*.
- [26] Y. Kalfoglou and M. Schorlemmer. Ontology mapping: the state of the art. *The Knowledge Engineering Review*, 18(1):1–31, January 2003.
- [27] A. Kalyanpur, B. Parsia, E. Sirin, and B.C. Grau. Repairing unsatisfiable concepts in OWL ontologies. In *ESWC*, pages 170–184. Springer, 2006.
- [28] Hirofumi Katsuno and Alberto O. Mendelzon. Propositional knowledge base revision and minimal change. *Artificial Intelligence*, 52(3):263–294, 1991.
- [29] I. Lakatos. *Proofs and Refutations: The Logic of Mathematical Discovery*. Cambridge University Press, 1976.
- [30] P. Langley. Descriptive discovery processes: Experiments in baconian science. Technical report, Carnegie-Mellon University, 1980.
- [31] P. Langley. The Computer-Aided Discovery of Scientific Knowledge. In *Discovery Science: First International Conference, DS’98, Fukuoka, Japan, December 14-16, 1998, Proceedings*. Springer, 1998.
- [32] D. B. Lenat. Automated theory formation in mathematics. In R. Reddy, editor, *Proceedings of IJCAI-77*, pages 833–842. International Joint Conference on Artificial Intelligence, August 1977.
- [33] D.B. Lenat. CYC: a large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11):33–38, 1995.
- [34] J.L. Mackie. *The cement of the universe: A study of causation*. Oxford: Clarendon Press, 1974.
- [35] J. Madhavan, P.A. Bernstein, P. Domingos, and A.Y. Halevy. Representing and Reasoning about Mappings between Domain Models. In *Proceedings of the National Conference on Artificial Intelligence*, pages 80–86. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2002.
- [36] F. Manola, E. Miller, et al. RDF Primer. *W3C Recommendation*, 2004.
- [37] P. McBrien and A. Poulouvasilis. Schema Evolution in Heterogeneous Database Architectures, A Schema Transformation Approach. *Proceedings of the 14th International Conference on Advanced Information Systems Engineering*, pages 484–499, 2002.

- [38] J. McCarthy. Circumscription — a form of non-monotonic reasoning. *Artificial Intelligence*, 13:27–39, 1980. Also in *Readings in Nonmonotonic Reasoning*, Ginsberg, M. L. (ed), Morgan Kauffman, 1987.
- [39] J. McCarthy. Applications of circumscription to formalizing common-sense knowledge. *Artificial Intelligence*, 28:89–116, 1986. Also in *Readings in Nonmonotonic Reasoning*, Ginsberg, M. L. (ed), Morgan Kauffman, 1987.
- [40] Deborah L. McGuinness and Frank van Harmelen. <http://www.w3.org/2001/sw/WebOnt/>, 2001.
- [41] F. McNeill and A. Bundy. Dynamic, automatic, first-order ontology repair by diagnosis of failed plan execution. *International Journal On Semantic Web and Information Systems*, 3(3):1–35, 2007. Special issue on ontology matching.
- [42] D. Miller and G. Nadathur. An overview of λ Prolog. In R. Bowen, editor, *Proceedings of the Fifth International Logic Programming Conference/Fifth Symposium on Logic Programming*. MIT Press, 1988.
- [43] I. Niles and A. Pease. Towards a standard upper ontology. In *Proceedings of the international conference on Formal Ontology in Information Systems*, pages 2–9. ACM Press, New York, NY, USA, 2001.
- [44] Tobias Nipkow, Lawrence C. Paulson, and Markus Wenzel. *Isabelle/HOL — A Proof Assistant for Higher-Order Logic*, volume 2283 of *LNCS*. Springer, 2002.
- [45] J.D. Norton. Causation as Folk Science. *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited*, 2007.
- [46] N.F. Noy and M. Klein. Ontology Evolution: Not the Same as Schema Evolution. *Knowledge and Information Systems*, 6(4):428–440, 2004.
- [47] Judea Pearl. *Causality : Models, Reasoning, and Inference*. Cambridge University Press, March 2000.
- [48] A. Pease, S. Colton, A. Smaill, and J. Lee. Modelling Lakatos’s philosophy of mathematics. In *Proceedings of the Second European Computing and Philosophy Conference, E-CAP2004*, University of Pavia, 2004.
- [49] P. Plessers and O. De Troyer. Ontology change detection using a version log. *Lecture notes in computer science*, pages 578–592.
- [50] G. Pólya. *How to Solve It*. Princeton University Press, 1945.
- [51] Ray Reiter. A logic for default reasoning. *Artificial Intelligence*, pages 81–132, 1980.
- [52] Raymond Reiter. On closed world data bases. In *Logic and Data Bases*, pages 55–76, 1978.
- [53] Alan Robinson and Andrei Voronkov, editors. *Handbook of Automated Reasoning*. Elsevier, 2001. 2 volumes.

- [54] J.F. Roddick. Schema evolution in database systems: an annotated bibliography. *ACM SIGMOD Record*, 21(4):35–40, 1992.
- [55] J. Rowe and D. Partridge. Creativity: a survey of AI approaches. *Artificial Intelligence Review*, 7(1):43–70, 1993.
- [56] V. C. Rubin, N. Thonnard, and Jr. Ford, W. K. Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from NGC 4605 / $R = 4\text{kpc}/$ to UGC 2885 / $R = 122\text{ kpc}/$. *Astrophysical Journal*, 238:471, 1980.
- [57] Bertrand Russell. On the notion of cause. In *Mysticism and logic, and other essays*, book IX, page 234 p. Longmans, Green, London :, 1918.
- [58] Nigel Shadbolt, Tim Berners-Lee, and Wendy Hall. The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101, 2006.
- [59] J.F. Sowa. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. MIT Press, 2000.
- [60] L. Stojanovic, A. Maedche, B. Motik, and N. Stojanovic. User-Driven Ontology Evolution Management. In *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web*, pages 285–300. Springer-Verlag London, UK, 2002.
- [61] L. Stojanovic, A. Maedche, N. Stojanovic, and R. Studer. Ontology evolution as reconfiguration-design problem solving. In *Proceedings of the 2nd international conference on Knowledge capture*, pages 162–171. ACM New York, NY, USA, 2003.
- [62] P. Thagard. *Coherence in Thought and Action*. MIT Press, 2002.
- [63] V. Ventrone. Semantic heterogeneity as a result of domain evolution. *ACM SIGMOD Record*, 20(4):16–20, 1991.
- [64] M. Wiser and S. Carey. When heat and temperature were one. In A. Stevens and D. Gentner, editors, *Mental Models*, pages 267–297. Erlbaum, 1983.
- [65] M. Wooldridge and N. R. Jennings. Intelligent agents: Theory and practice. *The Knowledge Engineering Review*, 10(2):115–152, 1995.
- [66] Michael Wooldridge. *An Introduction to MultiAgent Systems*. John Wiley and Sons, 2002.