Automating Signature Evolution in Logical Theories *

Alan Bundy

School of Informatics,
University of Edinburgh,
Edinburgh, UK
e-mail: A.Bundy@ed.ac.uk

1 Introduction

The automation of reasoning as deduction in logical theories is well established. Such logical theories are usually inherited from the literature or are built manually for a particular reasoning task. They are then regarded as fixed. We will argue that they should be regarded as fluid.

1. As Pólya and others have argued, appropriate representation is the key to successful problem solving [Pólya, 1945]. It follows that a successful problem solver must be able to choose or construct the representation best suited to solving the current problem. Some of the most seminal episodes in human problem solving required radical representational change.
2. Automated agents use logical theories called *ontologies*. For different agents to communicate they must align their ontologies. When a large, diverse and evolving community of autonomous agents are continually engaged in online negotiations, it is not practical to manually pre-align the ontologies of all agent pairs – it must be done dynamically and automatically.
3. Persistent agents must be able to cope with a changing world and changing goals. This requires evolving their ontologies as their problem solving task evolves. The W3C call this *ontology evolution*[1].

Furthermore, in evolving a logical theory, it is not always enough just to add or delete axioms, definitions, rules, etc. — a process usually called *belief revision*. Sometimes it is necessary to change the underlying signature of the theory, e.g., to add, remove or alter the functions, predicates, types, etc. of the theory.

Below we present two projects to automate signature evolution in logic theories: one in the domain of online agents and one in the domain of theories of physics. Common themes emerge from these two projects that offer hope for a general theory of signature evolution.

* The research reported in this paper was supported by EPSRC grant EP/E005713/1. It will soon be supported by EPSRC grant EP/G000700/1 I would like to thank Michael Chan, Lucas Dixon and Fiona McNeill for their feedback on this paper and their contributions to the research referred to in it.

[1] http://www.w3.org/TR/webont-req/\#goal-evolution

2 ORS: Diagnosing and Repairing Agent Ontologies

We first investigated the automation of signature evolution in ORS (Ontology Repair System): an automated system for repairing faulty ontologies in response to unexpected failures when executing multi-agent plans [McNeill & Bundy, 2007]. ORS forms plans to achieve its goals using the services provided by other agents. In forming these plans, ORS draws upon its knowledge base, which provides a representation of its world, including its beliefs about the abilities of other agents and under what circumstances they will perform various services. To request actions or ask questions of the other agents, ORS uses a simple performative language implemented in KIF[2], an ontology language based on first-order logic.

The representation of the world used by ORS may be faulty, not just in containing false beliefs, but also in using a signature that does not match that used by some of its collaborating agents. This mismatch will inhibit inter-agent communication, leading to faulty plans that will fail during execution. ORS analyses its failed plans, communicates with any agents that unexpectedly refused to perform a service, and proposes repairs to its ontology, including the signature of that ontology. Repairs can include: adding, removing or permuting arguments to predicates or functions, merging or splitting of predicates or functions and changing their types, as well as some belief revisions, such as adding or removing the precondition of an action.

Adding arguments to and splitting functions are examples of *refinement*, in which ontologies are enriched. Unfortunately, refinement operations are only partially defined. For instance, when an additional argument is added to a function it is not always clear what value each of its instances should take, or indeed whether any candidate values are available. When an old function is split into two or more new functions, each occurrence of the old function must be mapped to one of the new ones. It is not always clear how to perform this mapping.

The evaluation of ORS consisted of attempts to reproduce automatically the manual repairs we observed in KIF ontologies. Although this evaluation was successful, it was hampered by a lack of examples of before and after versions of ontologies, and of records of the fault in the before version, how it was diagnosed and how it was repaired to produce the after version. This led us to investigate domains in which ontological evolution was better documented. We picked the domains of physics and law. Our progress in the physics domain is the topic of the next section.

3 GALILEO: Signature Evolution in Physics

We are now applying and developing our techniques in the domain of physics [Bundy, 2007, Bundy & Chan, 2008]. This is an excellent domain because many of its most seminal advances can be seen as signature evolution, i.e., changing the way that physicists view the world. These changes are often triggered by

[2] `http://logic.stanford.edu/kif/kif.html`

a contradiction between existing theory and experimental observation. These contradictions, their diagnosis and the resulting repairs have usually been well documented by historians of science, providing us with a rich vein of case studies for the development and evaluation of our techniques, addressing the evaluation problem identified in the ORS project. The physics domain requires higher-order logic: both at the object-level, to describe things like planetary orbits and calculus, and at the meta-level, to describe the repair operations.

3.1 Repair Plans

We are developing a series of *repair plans* which operate simultaneously on a small set of small higher-order theories, e.g., one representing the current theory of physics, another representing a particular experimental set-up. Before the repair, these theories are individually consistent but collectively inconsistent. Afterwards the new theories are also collectively consistent. Each repair plan has a trigger formula and some actions: when the trigger is matched, the actions are performed. The actions modify the signatures and axioms of the old theories to produce new ones. Typical actions are similar to those described above for ORS. The repair plans have been implemented in the GALILEO system (Guided Analysis of Logical Inconsistencies Leads to Evolved Ontologies) using λProlog [Miller & Nadathur, 1988] as our implementation language, because it provides a polymorphic, higher-order logic.

This combination of repair plans and multiple interacting logic theories helps to solve several tough problems in automated signature evolution.

- The overall context of the plan completes the definition of the, otherwise only partially defined, refinement operations. For instance, it supplies the values of additional arguments and specifies which new function should replace which old one. Organising the theory as several interacting, small theories further guides the refinement, e.g., by enabling us to uniformly replace all the occurrences of an old function in one theory in one way, but all the occurrences in another theory in a different way.
- Grouping the operations into a predefined repair plan helps control search. This arises not only from inference, but also from repair choices. It also occurs not only within the evolving object-level theory but also in the meta-level theory required to diagnose and repair that object-level theory. This solution is adopted from our work on *proof plans* [Bundy, 1991].
- Having several theories helps us control inconsistency. A predictive theory and an observational one can be internally consistent, but inconsistent when combined. Since all sentences are theorems in an inconsistent theory, the triggers of all repair plans would be matched, creating a combinatorial explosion. This problem can be avoided when a trigger requires simultaneous matching across a small set of consistent ontologies.
- It is also enabling us to prove the minimality of our repair plans, i.e., to show that the repairs do not go beyond what is necessary to remove the inconsistency. We have extended the concept of *conservative extension* to

signature evolution. We can now prove that the evolution of each separate theory is conservative in this extended sense. Of course, we do not want the evolution of the combined theory to be conservative, since we want to turn an inconsistent combined theory into a consistent one.

3.2 Some Repair Plans and their Evaluation

We have so far developed two repair plans, which we call *Where's my stuff?* (WMS) and *Inconstancy*. These roughly correspond to the refinement operations of splitting a function and adding an argument, respectively. We have found multiple examples of these repairs across the history of physics, but are always interested in additional ones.

The WMS repair plan aims at resolving contradictions arising when the predicted value returned by a function does not match the observed value. This is modelled by having two theories, corresponding to the prediction and the observation, with different values for this function. To break the inconsistency, the conflicting function is split into three new functions: *visible*, *invisible* and *total*. The conflicting function becomes the total function in the predictive theory and the visible function in the observation theory[3]. The invisible function is defined as the difference between them, and this new definition is added to the predictive theory. The intuition behind this repair is that the discrepancy arose because the function was not being applied to the same *stuff* in the predictive and the observational theories — the invisible stuff was not observed.

WMS has been successfully applied to conflicts between predictions of and observations of the following functions: the temperature of freezing water; the energy of a bouncing ball; the graphs relating orbital velocity of stars to distance from the galactic centre in spiral galaxies; and the precession of the perihelion of Mercury. In these examples the role of the invisible stuff is played by: the latent heat of fusion, elastic potential energy, dark matter and an additional planet, respectively.

The Inconstancy repair plan is triggered when there is a conflict between the predicted independence and the observed dependence of a function on some parameter, i.e., the observed value of a function unexpectedly varies when it is predicted to remain constant. This generally requires several observational theories, each with different observed values of the function, as opposed to the one observational theory in the WMS plan. To effect the repair, the parameter causing the unexpected variation is first identified and a new definition for the conflicting function is created that includes this new parameter. The nature of the dependence is induced from the observations using curve-fitting techniques.

Inconstancy has been successfully applied to the following conflicts between predictions and various observations: the ratio of pressure and volume of a gas; and again the graphs relating orbital velocity of stars to distance from the galactic centre in spiral galaxies. The unexpected parameter of the function is the temperature of the gas and the acceleration between the stars, respectively. The

[3] There are situations in which these roles are inverted [Bundy, 2007]

first of these repairs generalises Boyle's Law to the Ideal Gas Law and the second generalises the Gravitational Constant to Milgrom's MOND (MOdified Newtonian Dynamics). Interestingly, WMS and Inconstancy produce the two main rival theories on the spiral galaxy anomaly, namely dark matter and MOND. Since this is still an active controversy, its unfolding will help us develop mechanisms to choose between rival theory repairs. We are currently experimenting also with applying Inconstancy to the replacement of Aristotle's concept of instantaneous light travel with a finite (but fast) light speed, using conflicts between the predicted and observed times of eclipses by Jupiter's moon Io.

4 Conclusion

We have argued for the importance of automated evolution of logical theories to adapt to new circumstances, to recover from failure and to make them better suited to the current problem. We argue that this requires more than just belief revision — although this is part of the story. We also need *signature* revision, i.e., changes to the underlying syntax of the theory. We have begun the work of automating signature evolution in the ORS and GALILEO projects. We have developed repair plans over multiple theories, which address some of the tough problems of partial definedness, combinatorial explosion, coping with inconsistency and ensuring minimality, that beset this endeavour.

In the future, we plan to: develop additional repair plans, research additional case studies from the history of physics, refine our currently rather *ad hoc* logical theories, thoroughly evaluate our repair plans on a significant corpus of case studies, and explore notions of minimal repair and other aspects of a theory of signature evolution. Ideas for future repair plans include: the converses of WMS and Inconstancy; the use of analogy to create new theories; and the correction of faulty causal dependencies.

References

[Bundy & Chan, 2008] Bundy, A. and Chan, M. (July 2008). Towards ontology evolution in physics. In Hodges, W., (ed.), *Procs. Wollic 2008*. LNCS, Springer-Verlag.

[Bundy, 1991] Bundy, A. (1991). A science of reasoning. In Lassez, J.-L. and Plotkin, G., (eds.), *Computational Logic: Essays in Honor of Alan Robinson*, pages 178–198. MIT Press.

[Bundy, 2007] Bundy, A. (July 2007). Where's my stuff? An ontology repair plan. In Ahrendt, W., Baumgartner, P. and de Nivelle, H., (eds.), *Proceedings of the Workshop on Disproving - Non-Theorems, Non-Validity, Non-Provability*, pages 2–12, Bremen, Germany. http://www.cs.chalmers.se/∼ahrendt/CADE07-ws-disproving/.

[McNeill & Bundy, 2007] McNeill, F. and Bundy, A. (2007). Dynamic, automatic, first-order ontology repair by diagnosis of failed plan execution. *IJSWIS*, 3(3):1–35. Special issue on ontology matching.

[Miller & Nadathur, 1988] Miller, D. and Nadathur, G. (1988). An overview of λProlog. In Bowen, R., (ed.), *Proceedings of the Fifth International Logic Programming Conference/ Fifth Symposium on Logic Programming*. MIT Press.

[Pólya, 1945] Pólya, G. (1945). *How to Solve It*. Princeton University Press.