Towards Ontology Evolution in Physics

Citation for published version:
Bundy, A & Chan, M 2008, Towards Ontology Evolution in Physics. in Logic, Language, Information and Computation: 15th International Workshop, WoLLIC 2008 Edinburgh, UK, July 1-4, 2008 Proceedings. vol. 5110, Lecture Notes in Computer Science, vol. 5110, Springer-Verlag GmbH, pp. 98-110. DOI: 10.1007/978-3-540-69937-8_10

Digital Object Identifier (DOI):
10.1007/978-3-540-69937-8_10

Published In:
Logic, Language, Information and Computation

Towards Ontology Evolution in Physics *

Alan Bundy and Michael Chan

School of Informatics,
University of Edinburgh,
Edinburgh, UK
{bundy, mchan}@inf.ed.ac.uk

Abstract. We investigate the problem of automatically repairing inconsistent ontologies. A repair is triggered when a contradiction is detected between the current theory and new experimental evidence. We are working in the domain of physics because it has good historical records of such contradictions and how they were resolved. We use these records to both develop and evaluate our techniques. To deal with problems of inferential search control and ambiguity in the atomic repair operations, we have developed *ontology repair plans*, which represent common patterns of repair. They first diagnose the inconsistency and then direct the resulting repair. Two such plans have been developed to repair ontologies that disagree over the value and the dependence of a function, respectively. We have implemented the repair plans in the GALILEO system and successfully evaluated GALILEO on a diverse range of examples from the history of physics.

1 Introduction

Most ontologies are built manually for a particular reasoning task. Successful reasoning depends on striking a compromise between the expressiveness of the representation and the efficiency of the reasoning process. If either the reasoning environment or the goals subsequently change, then the reasoning process is likely to fail because the ontology is no longer well suited to its task.

Many modern applications of automated reasoning need to work in a changing environment with changing goals. Their reasoning systems need to adapt to these changes automatically. In particular, their ontologies need to evolve automatically [1]. It is not enough to remove from or add to the beliefs of the ontology. It is necessary to change its underlying formal language. Our group has pioneered work in this new area of research. Our techniques involve diagnosis of faults in an existing ontology and then repairing these faults. They have previously been implemented and evaluated in the Ontology Repair System (ORS) [2].

We are now applying and developing our techniques in the domain of physics [3]. This is an excellent domain because many of its most seminal advances can be seen as ontology evolution, i.e., changing the way

* The research reported in this paper was supported by EPSRC grant EP/E005713/1.

that physicists view the world. These changes are often triggered by 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.

We will use the word 'ontology' generically, to refer to any logical theory, i.e., not restricted to ontologies in description logic, KIF or other logics in which ontologies have been traditionally formalised. 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 ontological repair operations. An ontology consists of two 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.

Our research will progress by composing together a number of diagnosis and repair operations into what we call *repair plans*, analysing a wide ranging development set of case studies, developing more repair plans and evaluating their performance on a test set of case studies. We will use this work as a basis to develop a theory of ontology evolution that we hope will also be applicable outwith the physics domain. We have already experimented with two repair plans, which we call "Where's my stuff?" (WMS) and "Inconstancy". In this paper we summarise our progress so far: describing our two repair plans, their implementation and their application to some seminal events in the history of physics.

2 Related Work

ORS evolved first-order ontologies by first diagnosing their faults via the execution failures of multi-agent plans, then using this diagnosis to guide repairs to these ontologies. These repairs were not just belief revisions, but changes to the underlying signatures, e.g., adding or removing function arguments, and splitting or conflating functions. These signature changes include but go beyond mere definitional extensions, i.e., adding definitions of new functions in terms of old ones. Some of the changes automated by both ORS and our current system extend the range of concepts that can be expressed.

There is also related work on repairing inconsistencies in OWL ontologies, for instance [6]. It focuses on strategies for removing inconsistent axioms and for identifying syntactical modelling errors in OWL ontologies to assist users to rewrite faulty axioms. Our focus, in contrast, is on repairing deeper conceptual errors in the underlying physical theory, rather than fixing errors in the use of the OWL operators. We are focusing on signature changes rather than removing or repairing axioms. We are also applying our techniques to higher-order rather than description logic ontologies. Nevertheless, we will be investigating this and other related work to identify opportunities for synergy.

More generally, our research differs from previous work on **ontology matching** by being focused on repairing a single ontology dynamically, automatically and without access to third party ontologies. It also differs from previous work on **belief revision** by being focused on signature changes rather than theory changes. However, again we will try to identify opportunities for synergy with these parallel strands of work.

3 Ontology Repair Plans

Adding arguments to and splitting functions are examples of *refinement*, in which ontologies are enriched. Such ontology refinement presents the following tough challenges:

1. 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 a function is split into two, it is not always clear to which of the new functions each occurrence of the old one should be mapped.
2. There are combinatorial explosions in both the object-level inference within the evolving ontology and the meta-level inference required to diagnose and repair that ontology.
3. To evaluate ORS we wanted to compare its evolutionary behaviour to that of manually executed ontology modifications. For this we needed a series of versions of such ontologies together with a justification of the modifications between successive versions. This proved very difficult to obtain: ontology developers do not usually keep publically accessible development histories[1]

The work outlined in this paper addresses these problems in the following way:

- Problems 1 and 2 are being addressed by developing *repair plans*. A repair plan is analogous to a *proof plan* [5], a generic proof outline that can be used to guide the search for a proof. Repair plans are generic combinations of diagnosis and repair operations that guide the ontology evolution process. By grouping these meta-level operations they trade off completeness against a dramatic reduction in search. Appendices A and B describe the two repair plans we have developed to date. In addition, we will develop a theory of ontology evolution by isolating and generalising the atomic ontology repair operations arising in our case studies. For instance, 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. Figure 1 outlines one approach we are currently exploring.
- Problem 3 is being addressed by working in the domain of physics. Some of the most seminal advances in the development of physics have required profound ontology evolution. Moreover, the evolutionary process in physics is very well documented. Detailed accounts

[1] A rare exception can be found in the CVS records at `http://sigmakee.cvs.sourceforge.net/sigmakee/KBs/`. Although, even here the annotations typically record *what* was changed but not much about *why*. We are also working on a project to re-evaluate ORS on some of these CVS records.

To define a concept of minimal repair we are experimenting with extending *conservative extension* to changes of signature, namely:

$$\phi \in Sig(O) \implies (\nu(O) \vdash \nu(\phi) \iff O \vdash \phi)$$

where $\phi \in Sig(O)$ means that ϕ is a formula in the signature of ontology O and $\nu(\phi)$ is the repaired ϕ in repaired ontology $\nu(O)$. In the repair plan in Appendix A on p11 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.

Fig. 1. Minimal Ontology Repair via Conservative Extensions

are available of: the problems with the prior ontology, e.g., a contradiction between theory and experiment; the new ontology; and an account of the reasoning which led to it.

The WMS ontology repair plan, described in detail in Appendix A, aims at resolving contradictions arising when the expected value returned by a function does not match the observed value. The expected value can be viewed as being deducible from an existing theory, whereas the observed value is obtained from some sensory data, which can be collected from empirical experiments. To break the inconsistency, the conflicting function in the theory is either split into two parts, *visible* and *invisible*, or becomes the visible part of a *total* function. The intuition behind this repair is that the discrepancy arose because the function was not being applied to the same *stuff* in the theory and the sensory ontologies - some of the stuff was invisible in one of the ontologies.

The Inconstancy ontology repair plan, described in Appendix B, is triggered when there is a conflict between the expected independence and the observed dependence of a function on some variable, i.e., the returned value of a function unexpectedly varies. To effect the repair, the variable causing the unexpected variation is first identified and a new, consistent definition for the conflicting function is then created. The new definition relies on a function relating the old definition to the varying condition. In our current implementation, this function is computed using curve fitting techniques.

4 Implementation

Our repair plans have been implemented in the *Guided Analysis of Logical Inconsistencies Leads to Evolved Ontologies* (GALILEO) system; we chose λProlog [4] 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 $-$,

$<$ and $=$. It also provides depth-first search to facilitate inference at both the object- and meta-levels. These features combines to support rapid prototyping of the repair plans and their application to the development case studies.
To illustrate the GALILEO code, the main clause of the WMS plan is provided in Appendix C.

5 Applications to the Development Case Studies

We have applied GALILEO to the emulation of a small but diverse set of physics case studies. For instance, WMS has been applied to the discovery of latent heat, an apparent paradox about the lost energy of a bouncing ball, the invention of dark matter and the speculation of the additional planet Vulcan to explain an anomaly in the precession of the perihelion of Mercury. Inconstancy has been applied to the extension of Boyle's Law to the Gas Law and the adaption of Newton's theory of gravity (MOND) to explain an anomaly in the rate of expansion of the Universe. §5.1 and §5.2 outline two of the applications of WMS and §5.3 outlines an application of Inconstancy.

5.1 The Application of WMS to the Discovery of Latent Heat

Joseph Black discovered the concept of latent heat around 1750. Wiser and Carey [7] 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:

$$O_t \vdash Heat(H_2O, Start(Freeze)) = Heat(H_2O, Start(Freeze)) \quad (1)$$

$$O_s \vdash Heat(H_2O, Start(Freeze)) = Heat(H_2O, End(Freeze)) \quad (2)$$

$$O_t \vdash Heat(H_2O, Start(Freeze)) \neq Heat(H_2O, End(Freeze)) \quad (3)$$

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. (1) comes from the reflexive law of equality, (2) comes from the observed constant temperature during freezing and (3) is deduced from the then current physical theory that heat decreases strictly monotonically when objects are cooled.
These formulae match the repair plan trigger (7) in Appendix A with the following substitution:

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

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 (8) in Appendix A 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 :

$$\nu(O_t) \vdash Heat(H_2O, Start(Freeze)) = Heat(H_2O, Start(Freeze))$$
$$\nu(O_s) \vdash Temp(H_2O, Start(Freeze)) = Temp(H_2O, End(Freeze))$$

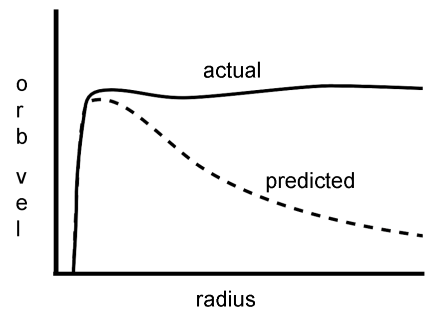
which breaks the derivation of the detected contradiction, as required.

5.2 The Application of WMS to Dark Matter

The evidence for dark matter arises comes from various sources, for instance, from an anomaly in the orbital velocities of stars in spiral galaxies identified by Rubin [8]. 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 2 illustrates 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 trigger the preconditions (7) in Appendix A with the following formulae:

$$O_t \vdash \lambda s \in Spiral.\ \langle Rad(s), Orb\_Vel(s) \rangle = Graph_p \quad (4)$$
$$O_s \vdash \lambda s \in Spiral.\ \langle Rad(s), Orb\_Vel(s) \rangle = Graph_a \quad (5)$$
$$O_t \vdash Graph_p \neq Graph_a \quad (6)$$

where $Orb\_Vel(s)$ 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 (4) shows the predicted graph, $Graph_p$: the orbital velocity decreases roughly inversely with the square root of the radius (see Figure 2). This graph is deduced by Newtonian Mechanics from the observed distribution of the visible



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.*

Fig. 2. Predicted *vs* Observed Stellar Orbital Velocities

stars in the spiral galaxy. Formula (5) shows the actual observed orbital velocity graph, $Graph_a$: it is almost a constant function over most of the values of s (see Figure 2). Note the use of λ abstraction in (4) and (5) to create graph objects as unary functions. These two graphs are unequal (6), within the range of legitimate experimental variation.

These three formulae instantiate the trigger preconditions (7) with the following substitution:

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

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 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 (8) suggested by this triggering is:

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

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:

$$\nu(O_t) \vdash \lambda s \in Spiral_{vis}.\ \langle Rad(s), Orb\_Vel(s)\rangle = Graph_p$$
$$\nu(O_s) \vdash \lambda s \in Spiral.\ \langle Rad(s), Orb\_Vel(s)\rangle = Graph_a$$

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 $Graph_a$.

5.3 The Application of Inconstancy to Modified Newtonian Mechanics

Another explanation of the anomaly in orbital velocities of stars in spiral galaxies is provided by MOdified Newtonian Dynamics (MOND), proposed by Moti Milgrom in 1981 as an alternative to the dark matter explanation. We have discussed in §5.2 that dark matter is an example of the WMS plan. MOND is an example of the Inconstancy plan. This is a good example of how the same observational discrepancies can trigger different repair plans. Essentially, MOND suggests that the gravitational constant is not a constant, but depends on the acceleration between the objects on which it is acting[2]. It is constant until the acceleration becomes very small and then it depends on this acceleration, which is the case for stars in spiral galaxies. So, the gravitational constant G can be repaired by giving it an additional argument to become $G(Acc(s))$, where $Acc(s)$ is the acceleration of a star s due to the gravitational attraction between the star and the galaxy in which it belongs. $Acc(s)$ is the variad and G is the inconstancy.

To satisfy the preconditions of the Inconstancy plan (10) and (10), we modify (4), (5), and (6) from §5.2. We want to have G instead of $\lambda s \in Spiral.\ \langle Rad(s), Orb\_Vel(s)\rangle$ on the left-hand-sides of (4) and (5). Similarly, we want to have the unrepaired representation of the gravitational constant on the right-hand-side of (4). We know from the law of universal gravitation that the square of the radius is inversely proportional to the acceleration of the orbiting star due to gravity, with the product of the gravitational constant and the mass of the galaxy being the constant of proportionality, i.e., $Rad(S_i)^2 = \frac{G \times M}{Acc(S_i)}$, where G, M, and $Acc(S_i)$ denote the gravitational constant, the mass of the galaxy, and the acceleration of the star w.r.t. the galaxy in which it belongs. So, we need to collect evidence for a variety of stars: S_i for $1 \leq i \leq n$, where $Acc(S_i)$ varies from large, i.e., S_i is near the centre of the galaxy, to small, i.e., S_i is on the periphery of the spiral galaxy. The graph shown in Figure 2, therefore, provides us with sufficient information to relate orbital velocities with accelerations of stars.

[2] It can also be presented as breaking of the equivalence of inertial and gravitational mass, but the varying gravity story fits our purposes better.

The trigger formulae for the Inconstancy plan will then be:

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

where $M2OV^{-1}$ is the inverse of $M2OV$, which takes the value of G, the mass of a star s $Mass(s)$ and the mass distribution of all the remaining stars in the spiral galaxy based on their positions $Posn(s)$, and calculates the orbital velocity of s. $M2OV^{-1}$, therefore, takes the observed orbital velocity of a star s $OV(s)$, the mass of s and the mass distribution of all the remaining stars in the galaxy and calculates what value of G would account for the observed orbital velocity of s.
The formulae above triggers the Inconstancy plan with the following substitution:

$$\{G/\mathit{stuff},\ \langle\rangle/\boldsymbol{s},\ \langle\rangle/\boldsymbol{x},\ 6.67 \times 10^{-11}/c,\ Acc/V,\ \langle S_i\rangle/\boldsymbol{b_i},\ G_1/c_1,\ G_n/c_n\}$$

Since G is a constant, both $\boldsymbol{s}$ and $\boldsymbol{x}$ are simply empty vectors.
Following the instructions for repair in Figure B, 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:

$$
\begin{aligned}
\nu(O_s(Acc(S_1) = A_1)) &\vdash \nu(G)(S_1) = M2OV^{-1}(OV(S_1), Mass(S_1),\\
&\qquad \lambda s \in \mathit{Spiral} \setminus \{S_1\}.\ (Posn(s), Mass(s)))\\
&\qquad (= G_1)\\
\vdots\ &\ \vdots\\
\nu(O_s(Acc(S_n) = A_n)) &\vdash \nu(G)(S_n) = M2OV^{-1}(OV(S_n), Mass(S_n),\\
&\qquad \lambda s \in \mathit{Spiral} \setminus \{S_n\}.\ (Posn(s), Mass(s)))\\
&\qquad (= G_n)\\
\nu(O_t) &\vdash \nu(G) ::= \lambda s.F(6.67 \times 10^{-11}, Acc(s))
\end{aligned}
$$

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.

6 Conclusion

Our proposed research programme is still in its early stages, although initial progress is promising. We have identified two of the five to ten general-purpose repair plans we seek, implemented them and applied them to the development set of case studies. Although small, this development set is satisfyingly diverse. Our current initial ontologies are *ad hoc*. We plan to develop more principled ones. Our prototype implementation requires the triggering formulae to be in just the right format, whereas later work will explore how to derive formulae meeting this format. This will raise difficult issues of search control. We need to evaluate our existing and future repair plans on a wider test set. We will require further investigation into the history of physics to identify both additional plans and both development and test case studies. Our theory of ontology evolution is in its infancy, although our extended notion of conservative extension is a promising start.

References

1. Bundy, A., McNeill, F.: Representation as a fluent: An AI challenge for the next half century. IEEE Intelligent Systems **21**(3) (2006) 85–87
2. McNeill, F., Bundy, A.: Dynamic, automatic, first-order ontology repair by diagnosis of failed plan execution. IJSWIS **3**(3) (2007) 1–35 Special issue on ontology matching.
3. Bundy, A.: Where's my stuff? An ontology repair plan. In Ahrendt, W., Baumgartner, P., de Nivelle, H., eds.: Proceedings of the Workshop on Disproving - Non-Theorems, Non-Validity, Non-Provability, Bremen, Germany (July 2007) 2–12 http://www.cs.chalmers.se/∼ahrendt/CADE07-ws-disproving/.
4. Miller, D., Nadathur, G.: An overview of λProlog. In Bowen, R., ed.: Proceedings of the Fifth International Logic Programming Conference/ Fifth Symposium on Logic Programming, MIT Press (1988)
5. Bundy, A.: A science of reasoning. In Lassez, J.L., Plotkin, G., eds.: Computational Logic: Essays in Honor of Alan Robinson, MIT Press (1991) 178–198
6. Kalyanpur, A., Parsia, B., Sirin, E., Grau, B.: Repairing unsatisfiable concepts in OWL ontologies. In: ESWC, Springer (2006) 170–184
7. Wiser, M., Carey, S.: When heat and temperature were one. In Stevens, A., Gentner, D., eds.: Mental Models. Erlbaum (1983) 267–297
8. Rubin, V.C., Thonnard, N., Ford, W. K., J.: Rotational properties of 21 SC galaxies with a large range of luminosities and radii, from

NGC 4605 /R = 4kpc/ to UGC 2885 /R = 122 kpc/. Astrophysical Journal **238** (1980) 471
9. Langley, P., Zytkow, J., Simon, H., Bradshaw, G.: Mechanisms for qualitative and quantitative discovery. In Michalski, R.S., ed.: Proceedings of the International Machine Learning Workshop, University of Illinois (June 1983) 121–132

A 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*[3] when it is applied to a vector of arguments $\boldsymbol{s}$ of type $\boldsymbol{\tau}$. $\mathit{stuff}(\boldsymbol{s})$ might, for instance, be the heat content of a block of ice or the orbit of a planet.

Trigger: If $\mathit{stuff}(\boldsymbol{s})$ 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.

$$O_t \vdash \mathit{stuff}(\boldsymbol{s}) = v_1,\, O_s \vdash \mathit{stuff}(\boldsymbol{s}) = v_2,\, O_t \vdash v_1 \neq v_2 \qquad (7)$$

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.

$$\forall \boldsymbol{s}{:}\boldsymbol{\tau}.\ \mathit{stuff}\,\sigma_{invis}(\boldsymbol{s}) ::= \mathit{stuff}(\boldsymbol{s}) - \mathit{stuff}\,\sigma_{vis}(\boldsymbol{s}) \qquad (8)$$

When *stuff* is a constant then σ_{vis}[4] 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:

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

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.

[3] *stuff* is a polymorphic, higher-order variable ranging over functions in physics.
[4] σ_{vis} and σ_{invis} are *replacements*. These resemble higher-order substitutions, except that constants, as well as variables, may be replaced with terms.

Note that the contradiction has now disappeared but the theorems of the two ontologies are preserved up to renaming and the logical consequences arising from adding the new *stuff* definition (8). Note also, that $=$, $>$ 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.

B The Inconstancy Ontology Repair Plan

Suppose that different sensory ontologies give distinct values for function $stuff(\boldsymbol{s})$ in different circumstances. Suppose function $V(\boldsymbol{s},\boldsymbol{b})$, where $\boldsymbol{b}$ contains variables distinguishing among these circumstances, returns distinct values in each of these circumstances, but is *not* one of the parameters in $\boldsymbol{s}$, i.e., $stuff(\boldsymbol{s})$ does not depend on $V(\boldsymbol{s},\boldsymbol{b})$. We will call $stuff(\boldsymbol{s})$ the *inconstancy* and $V(\boldsymbol{s},\boldsymbol{b})$ the *variad*. The Inconstancy repair plan establishes a relationship between the variad $V(\boldsymbol{b})$ and the inconstancy $stuff(\boldsymbol{s})$. 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 $stuff(\boldsymbol{s})$ is measured to take different values in different circumstances, then the following trigger formulae will be matched.

$$\begin{aligned} O_s(V(\boldsymbol{s},\boldsymbol{b_1}) = v_1) &\vdash stuff(\boldsymbol{s}) = c_1 \\ \vdots \quad &\quad \vdots \\ O_s(V(\boldsymbol{s},\boldsymbol{b_n}) = v_n) &\vdash stuff(\boldsymbol{s}) = c_n, \end{aligned} \tag{9}$$

$$O_t \vdash stuff(\boldsymbol{x}) ::= c(\boldsymbol{x}), \exists i \neq j. O_t \vdash c_i \neq c_j \tag{10}$$

where $\boldsymbol{x}$ can be instantiated by $\boldsymbol{s}$, $O_s(V(\boldsymbol{s},\boldsymbol{b_i}) = v_i)$ is the sensory ontology containing observations made under the condition that $V(\boldsymbol{s},\boldsymbol{b_i}) = v_i$ and $V(\boldsymbol{s},\boldsymbol{b})$ is not an existing argument of $stuff(\boldsymbol{s})$, i.e., $V(\boldsymbol{s},\boldsymbol{b}) \notin \boldsymbol{s}$.

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

$$\nu(stuff) ::= \lambda \boldsymbol{y},\boldsymbol{x}.\ F(c(\boldsymbol{x}), V(\boldsymbol{x},\boldsymbol{y})) \tag{11}$$

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:

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

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

To discover the meaning of the function F, we follow the traditional of Langley's BACON program [9] by using curve fitting. The $O_s(V(\boldsymbol{s}, \boldsymbol{b_i}) = v_i)$ ontologies provide a useful collection of equations: $F(c(\boldsymbol{s}), V(\boldsymbol{s}, \boldsymbol{b_i})) = c_i$ for $i = 1, \ldots, n$. Curve fitting techniques, e.g., regression analysis, are applied to these equations to approximate a definition of F. This hypothesis can then be tested by creating additional observations $O_s(V(\boldsymbol{s}, \boldsymbol{b_j}) = v_j)$, for new values of $V(\boldsymbol{s}, \boldsymbol{b_j})$, and confirming or refuting the hypothesis.

C Example Code

Here is the main clause[5] of the WMS plan.

```
repair O1 O2 NA1 NA2 :-
    % Repair triggered. Find stuff, args and parity
    wms_trigger O1 O2 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 O1 A1,
    % of both ontologies
    axioms O2 A2,
    % Flip to find opposite parity
    flip P FP,
    % Change both sets of axioms
    change P  O1 A1 STot SVis SInvis NA1,
    change FP O2 A2 STot SVis SInvis NA2.
```

`O1` and `O2` 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 (7) 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`.

[5] Confusingly, λProlog 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.