Where's My Stuff? An Ontology Repair Plan

Citation for published version:
Bundy, A 2007, Where's My Stuff? An Ontology Repair Plan. in Workshop on DISPROVING - Non-Theorems. vol. 4, CADE Inc.

Published In:
Workshop on DISPROVING - Non-Theorems

Where's My Stuff? An Ontology Repair Plan *

Alan Bundy

June 15, 2007

Abstract

Appropriate representation is the key to successful reasoning. Hence, if intelligent agents are to cope with changing goals in a changing environment, they must be able to adapt their representations, i.e., to detect that a current representation is inadequate, to diagnose its shortcomings and to repair it. In this paper we address the most basic kind of representational shortcoming: inconsistency. We focus on how certain kinds of inconsistency can be repaired using a *repair plan* that we entitle *Where's My Stuff?*. We apply this repair plan manually to four examples from the domain of Physics. In each case an inconsistent ontology is repaired into a consistent one. This extends the interest of the Disproving workshop beyond the "reparation of non-theorems" to the reparation of inconsistent ontologies. The Physics domain has the advantage that many faulty ontologies have been recorded by historians of science, together with the evidence that identified their faults and the ontological repairs that were proposed to mend them. These records provide plenty of data for developing and evaluating ontology repair plans.

1 Introduction

In [Bundy *et al*, 2006, McNeill & Bundy, forthcoming] we have described the Ontology Repair System (ORS), which repairs faulty, first-order ontologies by diagnosing the execution failures of multi-agent plans. These repairs were not just belief revisions, but changes to the underlying signatures, e.g., adding or removing predicate or function arguments, splitting or conflating predicates or functions. Adding arguments and splitting functions are examples of *refinement*, in which ontologies are enriched. An inherent problem with these refinement operations is that they are only partially defined. For instance, when an additional argument is added to a function it is not always clear what value each instance should take. 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.

In current work we are trying to develop a theory of ontology repair and to extend it to new domains. In this paper we report two advances.

- The aggregation of repair operations into *repair plans*, which helps address the partial definedness of the refinement operations.
- The development of the *Where's My Stuff*[1] repair plan and its manual application[2] to four examples from the Physics domain.

Our claim is that the *Where's My Stuff* plan can successfully account for the ontological repairs required in several historic advances in Physics. The evidence for this claim is the manually worked examples in the rest of this paper.

*The research reported in this paper was supported by EPSRC grant GR/S01771. Thanks to Graham Steel, three anonymous Disproving workshop referees and members of the *DREAM* Group for feedback on an earlier draft. Thanks to Alan Smaill for assistance with λProlog.

[1] With apologies to Amazon.

[2] An implementation in λProlog is under development.

Our immediate aim is the exploration of mechanisms for ontology repair. Physics is a convenient development domain due to the abundance of historical records of fault diagnosis and repair in Physical ontologies. Our objective is to build a prototype, possibly interactive, computer program which can emulate[3] a wide variety of these historical ontology repair episodes. In future, this work might lead to some practical application, but this is not our current objective.

By *repair plan* we mean a compound, possibly hierarchical, system of repair operations, with associated preconditions and effects, in the way that a proof plan [Bundy, 1991] is a compound system of rules of inference. The *Where's My Stuff* plan is intended to be the first move in gathering a portfolio of such repair plans, which we hope will collectively cover a large number of ontology repairs, at least in the Physics domain. For instance, we are currently developing a repair plan based on adding additional arguments to functions with unexpected variation. The triggering pattern for this new plan will overlap with that for *Where's My Stuff*, providing a rival ontology repair in some cases. Our hope is that just tens of similar repair plans will provide significant coverage of historical ontology repairs in Physics; we expect complete coverage to be impossible due to the need for idiosyncratic or domain-specific repairs in some cases.

Typed higher-order ontologies appear to be required in the Physics domain, since many of the concepts, for instance, calculus, are essentially higher-order and many of the functions only make sense when applied to objects of certain types, e.g., Orb_Vel takes an astronomical object and returns a real number. So, in this paper, *ontology* will mean a theory in typed, higher-order logic. We use the word "ontology" rather than "theory" because: (a) our emphasis on *signature* modification might be obscured by the word "theory"; (b) we intend, eventually, to apply the mechanisms we develop to ontology repair in the semantic web etc. and; (c) this frees up the word "theory" to refer to the set of *theorems* of the ontology. Higher-order logic is also required to describe the modifications to the functions of the ontology, i.e., to describe the mechanisms themselves.

An ontology O is a pair $\langle Sig(O), Ax(O)\rangle$, where $Sig(O)$ is the *signature* and $Ax(O)$ are the *axioms*. The *language* $L(O)$ of O is set of formulae generated by the grammar $Sig(O)$. The *theory* $Th(O)$ of O is the set of theorems generated from the closure of $Ax(O)$ over the rules of inference of simply-typed lambda calculus, e.g., expressed as a sequent calculus. We will write $O \vdash \phi$ when $\phi \in Th(O)$. If ontology O is faulty then the repaired ontology will be denoted $\nu(O)$, where ν is the function that converts the faulty ontologies into repaired ones, i.e., it is the instantiated repair plan. Since ontology repair could, in principle, involve changes to the underlying logic, we reserve the right to depart from these definitions (although not in this paper) and have deliberately left them a little open-ended. Note, in particular, the need for both a typed and higher-order logic, in contrast to the first-order, sorted logic of ORS. As argued above, these extensions are needed for the Physics domain.

We envisage these repair plans being implemented by higher-order deductive machinery. The preconditions of each repair plan will contain some patterns expressed as higher-order formulae. To trigger the repair plan, these patterns will be matched to the original ontology. When they match, then some higher-order output patterns will be instantiated and thereby form the repaired ontology. We have started to develop a prototype implementation in the higher-order logic programming language λProlog [Miller & Nadathur, 1988]. Since the typed, higher-order Physics ontologies required for this implementation do not already exist, we are constructing them on an 'as needed' basis for each of our test examples. Our examples below are intended to illustrate both the triggering and the repair mechanisms. The uniform presentation is intended to illustrate the essentially algorithmic nature of these processes, to convince the reader that they can be readily implemented. Human intervention is, however, currently required to prepare the ontology to facilitate the triggering process. There is a brief discussion of the deductive and search issues involved in such preparations, but the details are left to further work.

[3] I.e. "to rival with some degree of success", but not to provide a historically valid model.

2 The Where's My Stuff Ontology Repair Plan

The *Where's My Stuff* ontology repair plan is triggered by a mismatch between the predicted and the actual value of some Physics function on some object. Let us call this function *stuff*, a higher-order, variadic, function variable[4] from physical objects or systems to some values that can be added, usually the reals, but we will see, in §5 and §6, that there are other possibilities, so addition must be polymorphic. The predicted value is a deductive inference from the original ontology of Physics theory, say $O_t \vdash \mathit{stuff}(\vec{c}) = v_1$, i.e., v_1 is the value, predicted by the theoretical ontology O_t, when *stuff* is applied to the n arguments in the vector $\vec{c}$. However, the observed value is v_2, i.e., $O_s \vdash \mathit{stuff}(\vec{c}) = v_2$, where O_s is the sensory theory whose axioms record the experimental observations and whose theorems are deductions from these observations. The predicted and observed values differ, i.e., $v_1 \neq v_2$, resulting in a contradiction if the ontologies O_t and O_s are combined. So, to summarise, the triggering pattern is:

$$O_t \vdash \mathit{stuff}(\vec{c}) = v_1, \quad O_s \vdash \mathit{stuff}(\vec{c}) = v_2, \quad O_t \vdash v_1 \neq v_2 \tag{1}$$

Note that $=$ is polymorphic; it depends on the type of the value returned by *stuff*, e.g., reals. It also needs to be a bit fuzzy, since there is always some noise in experimental data. One way to achieve this would be to associate error bars with any value and count two values equal if the intersection of the intervals defined by these error bars was non-empty.

The repair is to split *stuff*, into three new terms: *stuff*, $\mathit{stuff}\,\sigma_{vis}$ and $\mathit{stuff}\,\sigma_{invis}$, where σ_{vis} and σ_{invis} are substitutions that replace one or more higher-order function variables with new functions of the same type. $\mathit{stuff}\,\sigma_{vis}$ and $\mathit{stuff}\,\sigma_{invis}$ are intended to be the visible and invisible parts of *stuff*, respectively. Then *stuff* is re-defined as the total of these two new functions:

$$\forall \vec{c}{:}\vec{\tau}.\ \mathit{stuff}(\vec{c}) \quad ::= \quad \mathit{stuff}\,\sigma_{vis}(\vec{c}) + \mathit{stuff}\,\sigma_{invis}(\vec{c}) \tag{2}$$

where the τ_i are the types of the c_i and $+$ is polymorphic, depending on the types of the values returned by *stuff*, e.g. reals. We have chosen to retain the name of the old *stuff* function as the name of the new total function. Alternatively, we could have retained it for the new visible part of *stuff* or we need not have retained it at all, choosing, say, a substitution σ_{total} to create a new term for the total function. These choices are just a matter of taste.

This new definition (2) is added to the axioms of the repaired theoretical ontology, $Ax(\nu(O_t))$. The remaining repairs to O_t and O_s depend on whether $v_1 > v_2$ or $v_1 < v_2$, where $>$ and $<$ are polymorphic total orders. If $v_1 > v_2$ then the remaining axioms of the repaired theory are copied unchanged[5] from the original axioms.

$$Ax(\nu(O_t)) \quad ::= \quad \{\phi \mid \phi \in Ax(O_t) \vee \phi = (2)\} \tag{3}$$

where $\phi = (2)$ is a convenient abuse of notation intended to abbreviate that ϕ might be the *stuff* definition axiom given in (2) above. (3) defines the axioms of the repaired ontology as being the old axioms plus (2). In particular, in the repaired ontology, $\nu(O_t) \vdash \mathit{stuff}(\vec{c}) = v_1$, i.e., the problematic prediction is preserved unchanged, but, as we shall see, it ceases to be problematic.

The sensory ontology O_s, however, *is* changed. We now take the observations of *stuff* to be observations of $\mathit{stuff}\,\sigma_{vis}$.

$$Ax(\nu(O_s)) \quad ::= \quad \{\phi\{\mathit{stuff}/\mathit{stuff}\,\sigma_{vis}\} \mid \phi \in Ax(O_s)\}$$

In particular, $O_s \vdash \mathit{stuff}\,\sigma_{vis}(\vec{c}) = v_2$. So, this observation no longer conflicts with the prediction. We can no longer directly observe values of *stuff*, but only of its visible part. At some future point, we hope we will devise methods to measure $\mathit{stuff}\,\sigma_{invis}$, but not at the current stage of repair.

[4]We use the anti-Prolog variable/constant convention: lower case letters are variables and upper case are constants.

[5]Since we have retained the old *stuff* name for the new total function.

If $v_1 > v_2$ then the roles of *stuff* and $stuff\sigma_{vis}$ are reversed: *stuff* is renamed to $stuff\sigma_{vis}$ in O_t but retained unchanged in O_s, i.e.,

$$Ax(\nu(O_t)) \ ::= \ \{\phi\{stuff/stuff\sigma_{vis}\} \mid \phi \in Ax(O_t) \vee \phi = (2)\}$$
$$Ax(\nu(O_s)) \ ::= \ \{\phi \mid \phi \in Ax(O_s)\}$$

So, the triggering formulae (1) are transformed to one of the following:

$$\nu(O_t) \vdash stuff(\vec{c}) = v_1, \quad \nu(O_s) \vdash stuff\sigma_{vis}(\vec{c}) = v_2, \quad if\ O_t \vdash v_1 > v_2 \tag{4}$$
$$\nu(O_t) \vdash stuff\sigma_{vis}(\vec{c}) = v_1, \quad \nu(O_s) \vdash stuff(\vec{c}) = v_2, \quad if\ O_t \vdash v_1 < v_2 \tag{5}$$

each of which breaks the previous derivation of a contradiction[6]. Note that this conditional branching on whether $v_1 > v_2$ ensures that $stuff\sigma_{invis}$ is always positive. Below, we will show examples of both (4) and (5).

Note how the *Where's My Stuff* repair plan overcomes the problem introduced in the last sentence of the first paragraph of §1: a function is split into three, but we are told exactly which occurrences of the original function turn into one of the new functions and which to leave unchanged. The repair also requires the addition of a new axiom and the mapping of some old derivations into new ones. Despite the compound structure and special properties of this repair plan, it is surprisingly widely applicable to the emulation of historical ontology repairs in Physics. In the next four sections we apply it, manually, to four such repairs, drawn from different areas of Physics and from different historical periods.

3 Application to the Latent-Heat Paradox

We start by applying it to the discovery of latent-heat by Joseph Black around 1750. [Wiser & Carey, 1983] discusses 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.

We can model this situation with the following formulae:

$$O_t \ \vdash \ Heat(H_2O, Start(Freeze)) = Heat(H_2O, Start(Freeze)) \tag{6}$$
$$O_s \ \vdash \ Heat(H_2O, Start(Freeze)) = Heat(H_2O, End(Freeze)) \tag{7}$$
$$O_t \ \vdash \ Heat(H_2O, Start(Freeze)) \neq Heat(H_2O, End(Freeze)) \tag{8}$$

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.

These formulae match the repair plan trigger (1) with the following substitution:

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

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) to:

$$\forall o{:}obj, t{:}mom.\ Heat(o,t) \ ::= \ Temp(o,t) + LHF(o,t)$$

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

[6] Which is not to say that some other contradiction does not still lurk undetected.

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 $Heat(H_2O, Start(Freeze)) > Heat(H_2O, End(Freeze))$, 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.

4 Application to the Bouncing-Ball Paradox

Our second example is based on an experiment described in [diSessa, 1983]. In [Bundy *et al*, 2006] we described it thus:

> "...consider the experiment conducted by Andreas diSessa on first-year MIT physics students [diSessa, 1983]. The students were asked to imagine a situation in which a ball is dropped from a height onto the floor. Initially, the ball has potential but not kinetic energy. Just before it hits the floor it has kinetic but not potential energy. As it hits the floor it has neither. Where did the energy go?"

The paradox arises because students typically idealise the ball as a particle without extent. However, the energy is stored in the compression of the ball[7] and this cannot be represented unless the idealisation of the ball has extent.

We can model this situation with the following formulae:

$$O_t \vdash TE(Ball, End(Drop)) = TE(Ball, Start(Drop)) \tag{9}$$

$$O_s \vdash TE(Ball, End(Drop)) = 0 \tag{10}$$

$$O_t \vdash TE(Ball, Start(Drop)) \neq 0 \tag{11}$$

where $Ball$ is the ball, $Drop$ is the time interval from its release to contact with the ground and $TE(Ball, t)$ is the total energy of the ball at time moment t. (9) comes from the law of conservation of energy; (10) comes from the observation that the ball is stationary and at zero height at the point of contact with the ground, so has neither potential nor kinetic energy; and (11) comes from the inference that the original energy of the ball consists of potential energy which is not zero. The substitution required to instantiate the trigger (1) with these three formulae is:

$$\{TE/stuff,\ \langle Ball, End(Drop)\rangle/\vec{c},\ TE(Ball, Start(Drop))/v_1,\ 0/v_2\}$$

To effect the repair we will define $\sigma_{vis} = \{TE_{part}/stuff\}$ and $\sigma_{invis} = \{EE/stuff\}$, so the new definition of TE that is proposed is:

$$\forall o{:}obj, t{:}mom.\ TE(o, t) ::= TE_{part}(o, t) + EE(o, t)$$

where $TE_{part}(o, t)$ is the total energy of a particle, defined as sum of its potential and kinetic energy, and $EE(o, t)$ is some invisible energy to be discovered. This invisible energy will turn out to be the elastic potential energy of the ball viewed as a spring: $EE(Ball, End(Drop))$. But the need to identify a source for this invisible energy could be the incentive to re-idealise the ball as an object with a type that has such an additional source of energy available, e.g., a spring.

In the repaired ontologies, since $TE(Ball, Start(Drop)) > 0$, the repaired triggering formulae are:

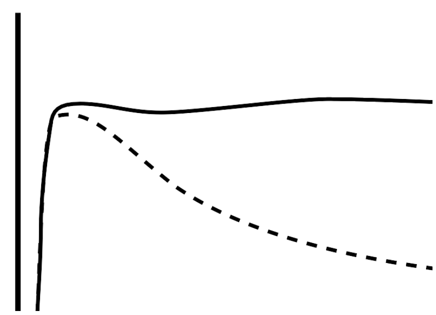
$$\begin{aligned} \nu(O_t) &\vdash TE(Ball, End(Drop)) = TE(Ball, Start(Drop)) \\ \nu(O_s) &\vdash TE_{part}(Ball, End(Drop)) = 0 \end{aligned}$$

which breaks the previous derivation of a contradiction, as required.

[7] And also of the floor, but we will ignore this factor in this exercise.

5 Application to Dark Matter

Our third example is the invention[8] of 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[9] identified by Rubin in 1975. 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 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.



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 observed graph.

Figure 1: Predicted *vs* Observed Stellar Orbital Velocities

We can trigger the preconditions (1) of the *Where's My Stuff* plan with the following formulae:

$$O_t \vdash \lambda s \in Spiral.\ \langle Rad(s), Orb\_Vel(s) \rangle = Graph_A \quad (12)$$

$$O_s \vdash \lambda s \in Spiral.\ \langle Rad(s), Orb\_Vel(s) \rangle = Graph_B \quad (13)$$

$$O_t \vdash Graph_A \neq Graph_B \quad (14)$$

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 (12) is the predicted orbital velocity graph based on the observed distribution of the visible stars and their masses in a spiral galaxy: the orbital velocity decreases inversely with the square root of the radius. Formula (13) is the observed orbital velocity graph: it is almost a constant function over most of the values of s. Note the use of λ abstraction to create graph objects as unary functions. These two graphs are unequal (14), within the range of legitimate experimental variation.

$Graph_A$ is deduced by Newtonian Mechanics from the observed distribution of mass in the spiral, i.e., it is a function, say, $M2OV$ (mass to orbital velocity) [10] of the mass distribution

[8] discovery?

[9] `http://en.wikipedia.org/wiki/Dark_matter`

[10] Alternatively, the actual formula might be inserted here, but it is enough for our purposes to know that such a formula exists, and the actual formula would clutter and obscure the picture. It involves complex calculus requiring computer calculation to give a solution.

graph: $\lambda s \in Spiral.\ \langle Rad(s), Mass(s)\rangle$:

$$O \vdash \lambda s \in Spiral.\ \langle Rad(s), Orb\_Vel(s)\rangle = M2OV(\lambda s \in Spiral.\ \langle Rad(s), Mass(s)\rangle) = Graph_A \quad (15)$$

Of course, the actual $M2OV$ calculation was done in the reverse direction: the mass distribution of the dark matter was calculated so that the predicted orbital velocities would fit the observational evidence.

These three formulae instantiate the trigger preconditions (1) with the following substitution:

$$\{\lambda s \in g.\ \langle Rad(s), Orb\_Vel(s)\rangle/stuff,\ \langle Spiral\rangle/\vec{c},\ Graph_A/v_1,\ Graph_B/v_2\}$$

Note that the repair plan works perfectly well with higher-order objects as the values v_1 and v_2, provided polymorphic $+$ and $\neq$ can be defined as having meaning over this data-type: in this case a piecewise addition 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 (2) suggested by this triggering is:

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

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. Note that $Spiral = Spiral_{vis} \cup Spiral_{invis}$, which is also an instantiation of definition (2), if you treat $\cup$ as addition for sets, but we cannot see how to trigger this simpler redefinition using the trigger (1).

In the repaired ontologies, since $Graph_A < Graph_B$, the repaired triggering formulae are:

$$\begin{aligned}\nu(O_t) &\vdash \lambda s \in Spiral_{vis}.\ \langle Rad(s), Orb\_Vel(s)\rangle = Graph_A \\ \nu(O_s) &\vdash \lambda s \in Spiral.\ \langle Rad(s), Orb\_Vel(s)\rangle = Graph_B\end{aligned}$$

which breaks the previous derivation of a contradiction, as required.

6 Application to the Precession of the Perihelion of Mercury

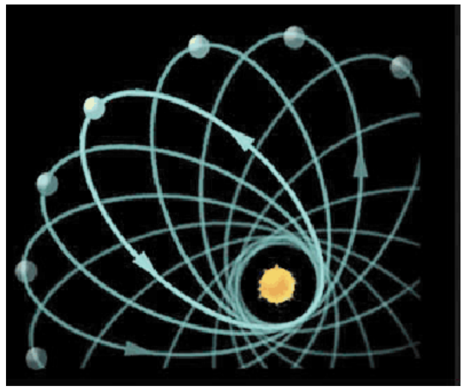
Our fourth, and last, example was suggested to us by the sociologist of science, Harry Collins: the precession of the perihelion of Mercury, i.e., the gradual rotation of the elliptical form of the orbit (see Figure 2). The orbits of the planets in the Solar System precess in this way. This is predicted by Newtonian Mechanics. However, Mercury's orbit does not precess by quite the right amount (`http://physics.ucr.edu/~wudka/Physics7/Notes_www/node98.html`). Nowadays, we understand this as an accurate prediction of Einstein's General Theory of Relativity, but for a long while it was believed to be caused by the gravitational attraction of an additional planet, named Vulcan, that was even closer to the Sun. Observation eventually ruled this out, but it is this (erroneous) prediction that we wish to model. Alternatively, we could have shown how a similar ontological repair could emulate the successful discovery of Pluto, but it is important to emphasise that our repair plan can be used to emulate ultimately *unsuccessful* ontology repairs, as well as successful ones.

We can represent the orbit of Mercury with the function $\lambda t.\ Posn(Mercury, t)$, where $Posn(o, t)$ is the 3D coordinate of object o at time t according to some implicit frame of reference. The triggering formulae are then:

$$O_t \vdash \lambda t.\ Posn(Mercury, t) = Orbit_p \quad (16)$$

$$O_s \vdash \lambda t.\ Posn(Mercury, t) = Orbit_o \quad (17)$$

$$O_t \vdash Orbit_p \neq Orbit_o \quad (18)$$



This diagram is taken from `http://physics.ucr.edu/~wudka/Physics7/Notes_www/node98.html#fig:prec.mercury`*. It shows the elliptical orbits of Mercury themselves rotating.*

Figure 2: Precession of the Perihelion of Mercury

where (16) is the predicted orbit of Mercury, (17) is the observed orbit and (18) asserts that these are not equal. Unfortunately, these triggers will not give us the right repair. What we would like is that *Solar_System* appeared in the term that instantiated *stuff* in (1). Then we could use (2) to define:

$$Solar\_System \quad ::= \quad Solar\_System_{vis} \cup Solar\_System_{invis}$$

where $Solar\_System_{invis} = \{Vulcan\}$. Unfortunately, we can't see a way of legitimately introducing *Solar_System* into the LHS of (16), say. However, all is not lost. The predicted orbit of Mercury, $Orbit_p$, is calculated by considering the mass distribution of the Solar System.

$$\begin{aligned} \lambda t.\ Posn(Mercury, t) &= M2O(\lambda s \in Solar\_System, t.\ \langle Posn(s,t), Mass(s)\rangle) \\ &= Orbit_p \end{aligned}$$

where $M2O$ is the calculation of the orbit of Mercury from the distribution of mass in the Solar System over time, i.e., taking into account the gravitational influences of the sun and the other planets. As with $M2OV$ in formula (15) in §5, $M2O$ will be some complex function involving calculus and requiring computer calculation, but for our purposes the details do not matter; it is enough that some such function exists.

By putting this calculation into reverse[11] ($M2O^{-1}$), we can create an alternative set of trigger formulae that *do* contain *Solar_System*, as required, namely:

$$\begin{aligned} O_t &\vdash \lambda o \in Solar\_System, t.\ \langle Posn(o,t), Mass(o)\rangle = M2O^{-1}(Orbit_p) \\ O_s &\vdash \lambda o \in Solar\_System, t.\ \langle Posn(o,t), Mass(o)\rangle = M2O^{-1}(Orbit_o) \\ O_t &\vdash M2O^{-1}(Orbit_p) \neq M2O^{-1}(Orbit_o) \end{aligned} \quad (19)$$

This set of triggers also has the nice property of predicting the mass distribution of the Solar System from the observed orbit of Mercury (this is what (19) means), and hence predicting the position of Vulcan. These three formulae instantiate the trigger preconditions (1) with the following substitution:

$$\{\lambda o \in s, t.\ \langle Posn(o,t), Mass(o)\rangle/stuff,\ \langle Solar\_System\rangle/\vec{c},\ M2O^{-1}(Orbit_p)/v_1,\ M2O^{-1}(Orbit_o)/v_2\}$$

[11] Assuming this *is* a function.

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

$$\begin{aligned} &\nu(O_t) \vdash \lambda o \in Solar\_System, t.\ \langle Posn(o,t), Mass(o)\rangle \\ &\quad = \lambda o \in Solar\_System_{vis}, t.\ \langle Posn(o,t), Mass(o)\rangle + \lambda o \in Solar\_System_{invis}, t.\ \langle Posn(o,t), Mass(o)\rangle \end{aligned}$$

as required. In the repaired ontologies, since $M2O^{-1}(Orbit_p) < M2O^{-1}(Orbit_o)$, the repaired triggering formulae are:

$$\begin{aligned} \nu(O_t) &\vdash \lambda o \in Solar\_System_{vis}, t.\ \langle Posn(o,t), Mass(o)\rangle = M2O^{-1}(Orbit_p) \\ \nu(O_s) &\vdash \lambda o \in Solar\_System, t.\ \langle Posn(o,t), Mass(o)\rangle = M2O^{-1}(Orbit_o) \end{aligned}$$

which breaks the previous derivation of a contradiction, as required.

This example demonstrates another dimension of choice: we may have to search back through the derivation of the contradicted prediction in order to find the most appropriate form for the trigger. We think this kind of choice was really also present in the previous examples, but we got lucky and managed to avoid making it explicit. For instance, in the Bouncing Ball example in §4, the trigger was phrased in terms of total energy, although this can hardly be directly observed. What can be observed is positions, distances and mass, from which the energies can be inferred and the contradiction derived. If this had been made more explicit then we think this dimension of search would have been revealed here too. Similar remarks can also be made about the dark matter example. Notice how this again forces us to mix theory and observation to deduce the theory-contradicting observation.

7 Conclusion

In this paper we have described the *Where's My Stuff* ontology repair plan. It consists of a higher-order, triggering pattern describing a particular kind of disagreement between theoretical predictions and experimental observations: a function *stuff* is predicted to have one value, but is observed to have another. When this pattern can be instantiated to a situation occurring in particular theoretical and sensory ontologies then this triggers some ontological repairs. The repairs have two main parts. Firstly, function refinement is applied to divide *stuff* into three functions: one for visible stuff, one for invisible stuff and one for their total. The original *stuff* is replaced by the total stuff in one ontology, but by the visible stuff in the other. Secondly, a new definition is added to the repaired theoretical ontology which defines the total function as the sum of the other two. These repairs disrupt the derivation of the contradiction.

We have applied this repair plan to four examples: latent heat, deSessa's bouncing ball , the invention of dark matter and the precession of the perihelion of Mercury. In each case we see that the repair plan provides a significant part of the required repair, but leaves some areas to be fixed. We also see that there are choices that require heuristic guidance. These choices arise in at least two ways: choices over which functions to refine into visible, invisible and their total, and choices over how far back to go in the derivations of the prediction and observation to identify the trigger formulae.

The dark matter and Mercury examples are particularly interesting, as *stuff* has to be instantiated to a compound λ term. It also shows the need for polymorphic $+$, $=$ and $>$, since these need to be interpreted differently depending on the data-type.

It's instructive to compare this repair plan with standard belief revision[12]. In belief revision the problem is to add some new belief ϕ to an existing ontology O. The interesting case is when just adding ϕ as a new axiom creates inconsistency. In this case, ϕ is usually assumed to take precedence over O, and O is adjusted by reducing its theory, up to and including removing O altogether. All this is done within a fixed signature. The *Where's My Stuff* repair plan, is triggered because an inconsistency will be created if we add, say, an observation ϕ to ontology O_t.

[12] http://en.wikipedia.org/wiki/Belief_revision

But ϕ does not take precedence. In fact, the signature is changed [13] so that *both* ϕ and O_t can be retained, but with a modified understanding of what each means. One is now seen to apply to a larger totality of stuff than previously assumed and the other just to the original, narrower conception of visible stuff. In this way, the inconsistency is made to melt away, but we come away with a richer conception of the world's complexity, including new questions about how to investigate the newly hypothesised invisible stuff.

Implementing a repair mechanism based on this repair plan requires higher-order matching and deduction, as well as some search control. The higher-order logic-programming language λProlog is well suited as an implementation, since it embodies all three elements. A λProlog implementation is currently under development.

We've been surprised to discover just how general the *Where's My Stuff* plan is. When we started this work we thought we knew a couple of examples, and had rejected two others: the bouncing ball and the precession of Mercury's orbit. But these rejected ones turned out to be examples too. Generality is just what we want, but, of course, it won't be enough. For instance, if it is applied too often we will get 'epicycles'. When we have 10 kinds of matter ranging from very light to very dark, then we will know it is time to apply the Occam's Razor repair plan $\smile$. We wonder what other plans are out there waiting to be revealed.

References

[Bundy, 1991] Bundy, Alan. (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. Also available from Edinburgh as DAI Research Paper 445.

[Bundy *et al*, 2006] Bundy, A., McNeill, F. and Walton, C. (2006). On repairing reasoning reversals via representational refinements. In *Proceedings of the 19th International FLAIRS Conference*, pages 3–12. AAAI Press. Invited talk.

[diSessa, 1983] diSessa, A. (1983). Phenomenology and the evolution of intuition. In Stevens, A. and Gentner, D., (eds.), *Mental Models*, pages 15–33. Erlbaum.

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

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

[Wiser & Carey, 1983] Wiser, M. and Carey, S. (1983). When heat and temperature were one. In Stevens, A. and Gentner, D., (eds.), *Mental Models*, pages 267–297. Erlbaum.

[13]To the best of our knowledge, our use of *signature* changes to affect ontological repairs was first introduced in our earlier ORS work. In particular, signature change has not been used in belief revision.