# Theory Evolution in Physics

Alan Bundy
(Joint work with Jos Lehmann and Michael Chan)

School of Informatics,
University of Edinburgh

University of St Andrews, $27^{th}$ November 2013

# Outline



# Representational Change in Scientific Revolutions

Concepts Splitting: heat becomes energy and temperature.

*one function becomes two.*

Concepts Merging: morning star and evening star become Venus.

*two constants become one.*

Dependencies Increase: gravitational 'constant' depends on accel.

*function gains an argument.*

Dependencies Decrease: gravitational accel independent of mass.

*function loses an argument.*

# Evolution in 'Normal' Science

- "Essentially, all models are wrong, but some are useful" (Box & Draper 1987)

  "But, it's turtles all the way down." (old lady quoted by Hawking 1988)
- Deciding what to include and what to neglect:

  *frictionless pulleys, inextensible strings, inelastic collisions.*
- Experimental evidence may cause initial modelling decision to be revoked.

# Issues in Theory Evolution

- Need theory development history for evaluation.
  - Lots of well documented case studies in physics domain.
- Search arises from diagnosis/repair choices as well as inference.
  - *Proof plans* previously used to control search.
  - Adapted proof plans to *theory repair plans*.
- Representational refinements incompletely defined.
  - Splitting a function: which new function replaces each occurrence of an old one?
  - Adding an argument: what values should its occurrences take?
  - Theory repair plans complete refinement definitions.
  - Multiple small theories facilitate targeted repairs.

# Theory Repair Plans

Aggregate atomic operations into theory repair plans.

- Atomic operations: split/merge functions, add/remove arguments, etc.
- Trigger formula signals applicability and instantiates plan.
- Repairs change language as well as beliefs.
- Multiple theories used in both trigger and repair.

# Example Theory Repair Plans

- The *Where's My Stuff?* theory repair plan.
  - Triggered by conflict between predicted and observed values of *stuff*.
  - Splits old *stuff* into visible, invisible and total *stuff*.
- The *Inconstancy* theory repair plan.
  - Triggered by conflict between predicted independence and observed dependence.
  - Adds this dependency as new argument to *stuff*.
- The *Unite* theory repair plan.
  - Triggered because two objects are the same on defining property.
  - Equates these two objects.

# Why Higher-Order Logic?

- Many concepts in Physics are naturally higher-order,
  - e.g. orbit of planet: $\lambda t.\ \mathit{Posn}(\mathit{Uranus}, t)$.
  - differential and integral calculus: $\mathit{Diff}(\mathit{Sin}) = \mathit{Cos}$.
  - So, higher-order needed at object-level.
- Polymorphic nature of theory repair plans.
  - e.g. $\mathit{stuff}$, $=$, $<$ $-$.
  - Need to quantify over types.
  - So, higher-order needed at meta-level too.

# The GALILEO System

- Implementing TRPs in the GALILEO system using Isabelle and its locales.

  GALILEO: *Guided Analyses of Logical Inconsistencies Leads to Evolved Theories.*

  Isabelle: Generic proof assistant. Higher-order theorem prover.

  Locales: Isabelle's mechanism for context handling.

- Successfully evaluated on several TRPs.

# The 'Where's My Stuff?' Theory Repair Plan

## Trigger

$$O_t \vdash \mathit{stuff}(\vec{s}) = v_1, \quad O_s \vdash \mathit{stuff}(\vec{s}) = v_2, \quad O_t \vdash v_1 > v_2$$

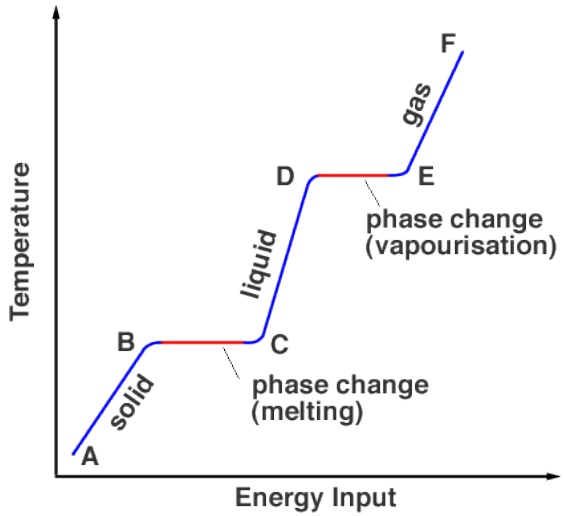
## Split stuff

$$\forall \vec{s} : \vec{\tau}.\ \mathit{stuff}_{\mathit{invis}}(\vec{s}) ::= \mathit{stuff}(\vec{s}) - \mathit{stuff}_{\mathit{vis}}(\vec{s})$$

## Create new axioms

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

# The Latent-Heat Paradox


- Before Joseph Black's investigations, heat and temperature conflated.
- Leads to paradox when heat is reduced but temperature is constant!
- Latent heat: the (hidden) heat energy involved in the phase change of a substance.

# Application to the Latent-Heat Paradox

## Example

**Original theories:**

$$\begin{aligned} O_t &\vdash Heat(H_2O, Melt) = Heat(H_2O, Melt) \\ O_s &\vdash Heat(H_2O, Melt) = 0 \\ O_t &\vdash Heat(H_2O, Melt) > 0 \end{aligned}$$

---

**Substitution:** $\{Heat/stuff,\ \langle H_2O, Melt\rangle/\vec{s},\ Heat(H_2O, Melt)/v_1, 0/v_2\}$

---

**Splitting stuff:**

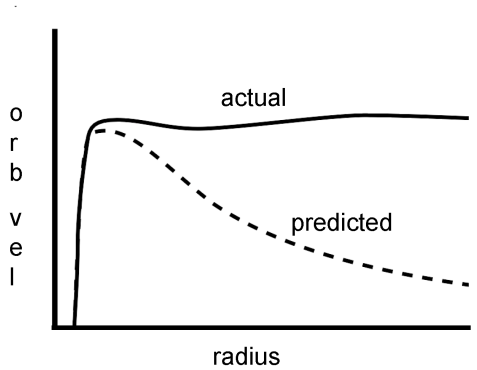
$\forall o : obj, e : event.\ LHF(o, e) ::= Heat(o, e) - Temp(o, e)$

---

**Repaired theories:**

$$\begin{aligned} \nu(O_t) &\vdash Heat(H_2O, Melt) > 0 \\ \nu(O_s) &\vdash Temp(H_2O, Melt) = 0 \end{aligned}$$

# Dark Matter




- Newtonian mechanics can predict relationship between orbital velocity and radius of stars.
- But actual relationship is different!
- Dark matter: some invisible matter surrounding the visible matter in a halo.

# Application to the Discovery of Dark Matter

## Example

**Original theories:**

$$\begin{aligned} O_t &\vdash \lambda s \in \mathit{Spiral}.\ \langle \mathit{Rad}(s), \mathit{Orb\_Vel}(s)\rangle = \mathit{Graph}_p \\ O_s &\vdash \lambda s \in \mathit{Spiral}.\ \langle \mathit{Rad}(s), \mathit{Orb\_Vel}(s)\rangle = \mathit{Graph}_a \\ O_t &\vdash \mathit{Graph}_p < \mathit{Graph}_a \end{aligned}$$

---

**Substitution:** $\{\lambda s \in g.\ \langle \mathit{Rad}(s), \mathit{Orb\_Vel}(s)\rangle / \mathit{stuff},\ \langle \mathit{Spiral}\rangle / \vec{s},\ \mathit{Graph}_p / v_1, \mathit{Graph}_a / v_2\}$

---

**Splitting stuff:**

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

---

**Repaired theories:**

$$\begin{aligned} \nu(O_t) &\vdash \lambda s \in \mathit{Spiral}_{vis}.\ \langle \mathit{Rad}(s), \mathit{Orb\_Vel}(s)\rangle = \mathit{Graph}_p \\ \nu(O_s) &\vdash \lambda s \in \mathit{Spiral}.\ \langle \mathit{Rad}(s), \mathit{Orb\_Vel}(s)\rangle = \mathit{Graph}_a \end{aligned}$$

# The Inconstancy Theory Repair Plan

## Trigger

$$\begin{array}{rcl} O_t & \vdash & stuff(\vec{x}) ::= c(\vec{x}) \\ O_s(V(\vec{s}, \vec{b_1}) = v_1) & \vdash & stuff(\vec{s}) = c_1(\vec{s}) \\ \vdots & \vdots & \vdots \\ O_s(V(\vec{s}, \vec{b_n}) = v_n) & \vdash & stuff(\vec{s}) = c_n(\vec{s}) \\ \exists i \neq j . O_t & \vdash & c_i(\vec{s}) \neq c_j(\vec{s}) \end{array}$$

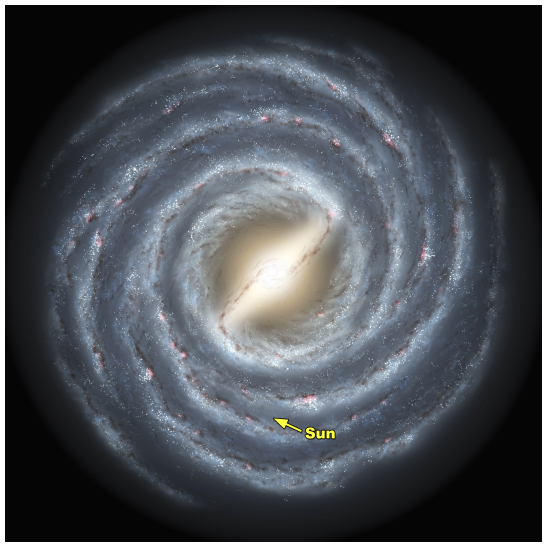
## Add variad

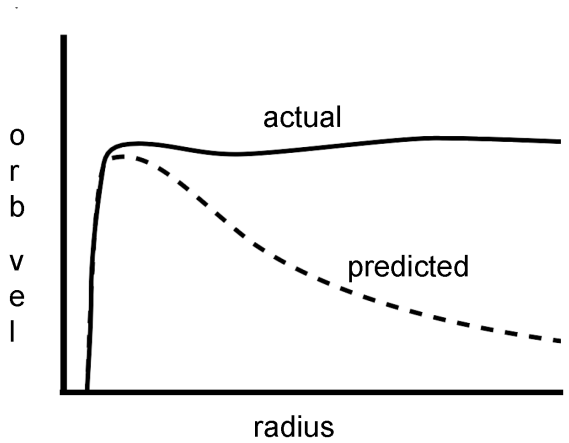
$$\nu(stuff) ::= \lambda \vec{y}, \vec{x}.\ F(c(\vec{x}), V(\vec{x}, \vec{y}))$$

## Create new axioms

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

# MOdified Newtonian Dynamics (MOND)




- Newtonian theory of gravity predicted that objects further out will have lower velocities.
- The observed velocities of those objects are almost constant!
- MOND - The gravitational force ($F = \frac{G \times M \times m}{r^2}$) is different at low accelerations.

# Application to MOND

## Example

**Original theories:**

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

---

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

---

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

---

**Repaired theories:**

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

# The Unite Theory Repair Plan

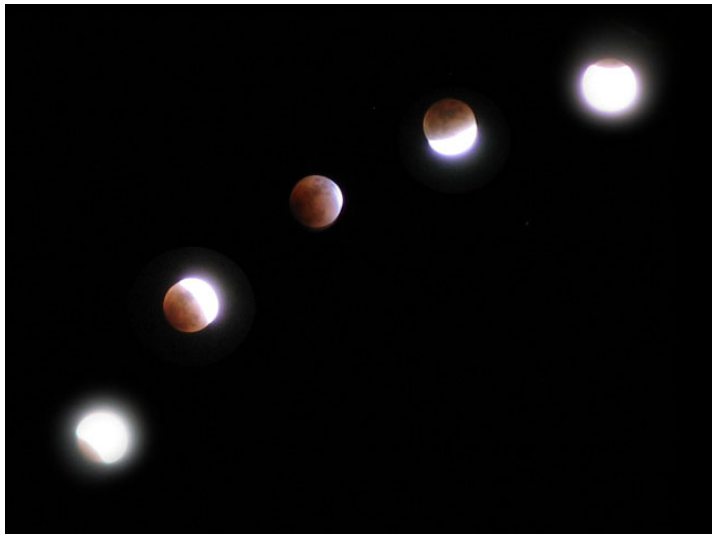
Trigger:

$$\begin{aligned} O_t &\not\vdash stuff_1 = stuff_2, \\ O_t &\vdash stuff_1{:}\tau \wedge stuff_2{:}\tau \wedge DefProp(dp, \tau), \\ O_s &\vdash dp(stuff_1) = dp(stuff_2). \end{aligned}$$

Repair:

$$Ax(\nu(O_t)) \quad ::= \quad \{stuff_1 = stuff_2\} \cup Ax(O_t)$$

# The Earth as a Sphere



- Shadow of Earth on Moon always circular.
  - Seen during lunar eclipse.
- Pythagoras realised family of projections was *defining property* of volumes.
- Only sphere has only circular projections.
- Therefore, Earth is spherical.

# Application to the Earth as a Sphere

Trigger:

$$
\begin{array}{lll}
O_t & \not\vdash & Shape(Earth) = Shape(Ball) \\
O_t & \vdash & Shape(Earth) : vol \wedge Shape(Ball) : vol \\
 & & \quad \wedge DefProp(\lambda v, t.\ project(v, Sun, Moon, t), vol) \\
O_s & \vdash & \lambda t.\ project(Shape(Earth), Sun, Moon, t) \\
 & & \quad = \lambda t.\ project(Shape(Ball), Sun, Moon, t)
\end{array}
$$

Repair:

$$Ax(\nu(O_t)) \quad ::= \quad \{Shape(Earth) = Shape(Ball)\} \cup Ax(O_t)$$

# Conservative Extensions and Minimal Repairs

- Want repairs to be minimal.
- Adapt *conservative extension.*

$$\phi \in \mathit{Lang}(O) \implies (\nu(O) \vdash \nu(\phi) \iff O \vdash \phi)$$

- In WMS, both $\nu(O_t)$ and $\nu(O_s)$ are conservative in this sense.

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

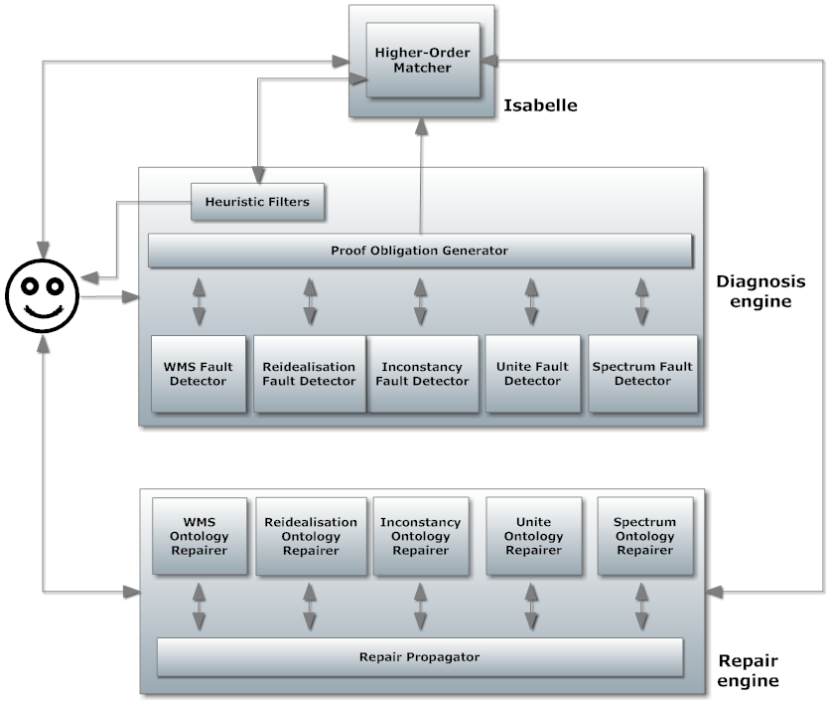
- The combined theories are *not* conservative.
- Situation more complicated for Unite and Inconstancy.

# Multiple Small Theories in Repair Plans

Why have separate theoretical and sensory theories?

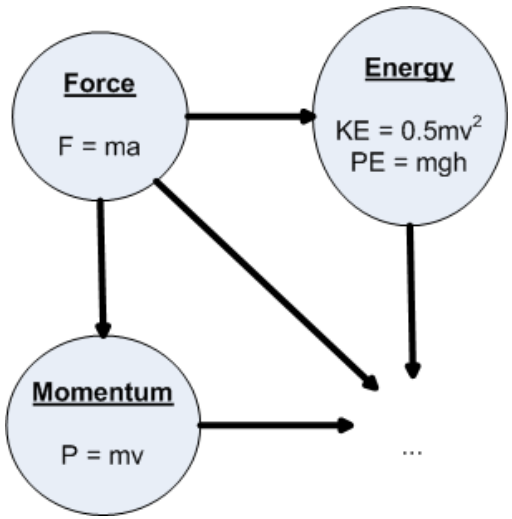
- Enables control over contradiction.
- Focuses effect of repair operations.
- Allows use of conservative extension to show minimality.
- Could allocate different degrees of belief to each theory.

# Overview of Architecture




- The reasoning engine tries to prove the repair trigger.
- If this succeeds, then the transformation rules are applied to the development graph.

# Development Graphs in Isabelle Locales




- Based on Institution Theory:
  - Nodes are theories; arrows are morphisms.
- Represent multiple theories in development graph
- Later theories can build on earlier ones.
- Add/remove theories automatically during evolution.

# Inferring Trigger Formulae in Isabelle

- GALILEO coordinates Isabelle's proof of trigger formula.
- Isabelle's locales mechanism used to organise multiple theories.
- Proofs interactive.
- *stuff*, *f*, etc. instantiated during proof.
- Repairs then suggested by theory repair plan.

# Search Control Heuristics

- Vast number of potential instantiations of $f$ and *stuff*.
- Heuristically prune instantiations with no Physics meaning, e.g., for *stuff*, filter out instantiations:
  - containing the identity function: $\ldots \lambda x.\, x \ldots$;
  - containing unbound function variables: $\ldots x(\ldots) \ldots$;
  - not containing a constant, e.g., $x(y, z)$;
  - with a $\lambda$ abstracted variable as head: $\lambda x.\, x(\ldots)$.

  Similar for $f$.
- Usually remaining instantiations suggest a small number of physical meaningful alternative repairs.

# Conclusion

- Theory evolution is a key technology for adaptive, autonomous agents.
- Physics is good development domain because of historical record.
- Higher-order logic needed at object- and meta-levels.
- Repair plans address problems of search and ambiguity.
- Developed and tested several repair plans.
- Implemented GALILEO system using Isabelle and its locales.