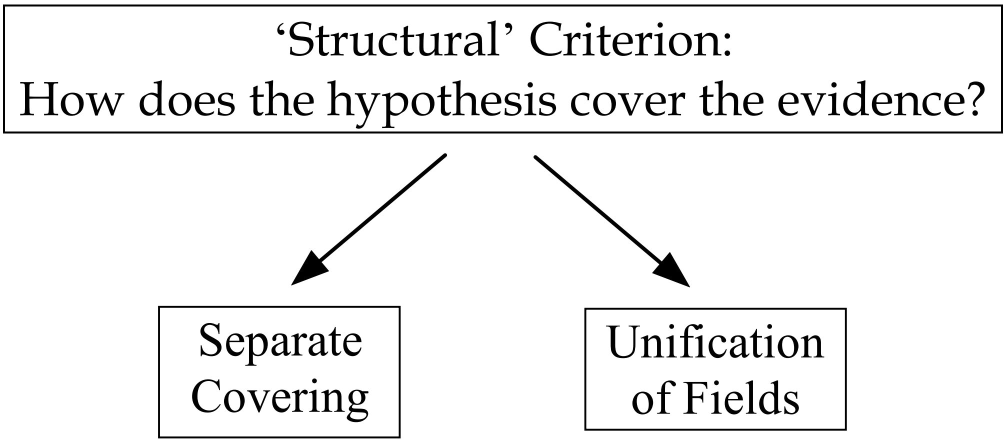
Computational ‘Consilience’ as a Basis for Theory Formation

José Hernández-Orallo
Universitat Politècnica de València
Departament de Sistemes Informàtics i Computació
Camí de Vera s/n, E-46022, València, Spain.
E-mail: `jorallo@dsic.upv.es`.

MODEL-BASED REASONING IN SCIENTIFIC DISCOVERY (MBR'98)
Pavia, Italy, December 17-19, 1998

What is ‘Consilience’?

Term coined in 1847 by Whewell for the selection of inductive theories:



Consilience: the evidence is ‘conciliated’ or unified by the theory.

Related concepts:

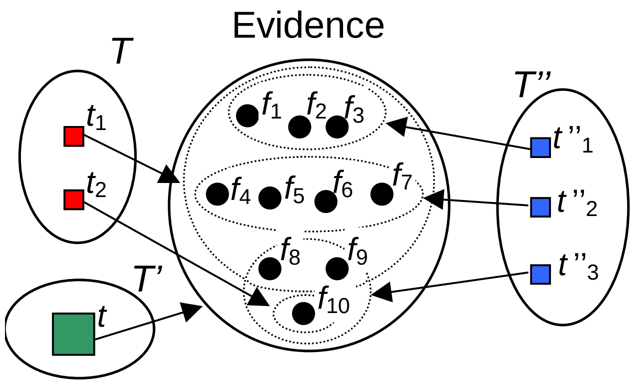
- Principle of ‘Common Cause’ (Reichenbach 1956).
- Explanatory Induction and Scientific Explanation (Harman 1965, Hempel 1965, Ernis 1968).
- Coherence (Thagard 1978).
- Intensionality *vs.* Tolerance of Partial Extensionality.

Distinguishing Consilience

EXAMPLE:

- Evidence $E = \{ f_1, f_2, \ldots f_{10} \}$
- Hypotheses:

  $T = \{ t_1, t_2 \}$, $T' = \{t''\}$ and $T'' = \{t'_1, t'_2, t'_3\}$



T'' shows two different (but closely related) notions:

- *T'* and *T''* are *intensional*. They have no exceptions.
- *T* and *T''* are separable. They are not *consilient*.

Towards Computational Consilience

We denote $M^+(T)$ the model of a theory T.

DEFINITION 1. **Separable Theories**
A theory T is n-separable in the partition of *different* theories $\Pi = \{ T1, T2, \ldots, Tn \}$ iff

$$M^+(T) = \cup_{i=1..n} M^+(T_i)$$
and
$$\forall_{i=1..n} \quad M^+(T_i) \neq \varnothing$$

Additional restrictions (*modes*) of separation:

I. *non-empty*: DEF 1
II. *non-subset*: DEF 1 $+\forall_{i,j=1..n} (P_i \subseteq P_j \Rightarrow i=j)$.
III. *disjoint*: DEF 1 $+ \forall_{i,j=1..n} (Pi \cap Pj = \varnothing)$.
IV. *non-subset model*: DEF 1 $+\forall_{i,j=1..n} (M^+(Pi) \subseteq M^+(Pj) \Rightarrow i=j)$.
V. *disjoint model*: DEF 1 $+\forall_{i,j=1..n} (M^+(Pi) \cap M^+(Pj) = \varnothing)$.

A theory is consilient iff it is not separable.

- The modes give 5 characterisations of consilient theories.

Example

EXAMPLE (Using Horn Theories):

- P_1= { p(a). q(X) :- r(X). r(a). } is {i-v}separable into Π = {{p(a)} , {q(X) :- r(X). r(a)}}.

- P_2= { q(X) :- r(X). r(b). } is not {ii-v}separable.

- P_3= { q(X) :- r(X). p(X) :- r(X). r(a). } is non-subset (model) separable into Π = {{ q(X) :- r(X). r(a)}, {p(X) :- r(X). r(a). }} but it is not disjoint (model) separable.

- P_4= { q(a). p(X) :- q(X). p(a) } is non-subset (model) and disjoint separable into Π = {{ q(a). p(X) :- q(X). }, {p(a).}} but it is not disjoint model separable.
  But there is Π′ = {{ q(a). } {p(X) :- q(X). p(a). }}

- P_5= { s(X):- p(X), q(b). p(X) :- q(X). t(X):-p(X),q(a) } is non-subset (model) and disjoint separable model into Π = {{ s(X) :- p(X), q(b). p(X) :- q(X) }, { p(X) :- q(X), t(X) :- p(X), q(a) } but it is not disjoint separable.

Problems of non-modularity (I, II, IV):

Problems of fantastic consilient concepts (III, IV & V):
P_1 can be '*conciliated*' by a fantastic concept *f*
into P'_1= { p(a) :- *f*. q(X) :- r(X), *f*. r(a), *f*. *f*. } for iii-iv.

Exception-Free Theories

DEFINITION 2. **Exceptions in a Theory**
A theory T has e = card($M^+(T_E)$) c-exceptions generated from T_E, denoted $\Delta_C(T, T_E) = e$, iff there is a partition $T = \{ T_R, T_E \}$ such that:

$$l(T) - l(T_R) \geq (1 / c) \cdot [l(M^+(T)) - l(M^+(T_R))]$$

where l denotes any syntactical measure of length.

If T_E is not specified, $\Delta_C(T) = \max \{ e \mid \Delta_C(T, T_E)=e \}$
Fixing l and an exception-degree c (usually c = 1), a theory T is said to be exception-free iff $\Delta_C(T) = 0$

Pragmatics:

- The modes give 5 characterisations of intensional (exception-free) theories.
- Mode ii and c=1 allow modular programs and avoid fantastic concepts.
- For instance, T = { p(X). q(X) } for evidence {p(a), p(b), p(e), q(a), q(d), q(e), q(f)} is separable but it has no exceptions.

Reinforcement (1 of 2)

Further detail on the relation hypothesis $\leq$ *evidence*:

Given a theory, a rule or component r_i is necessary for e iff $T \models e \quad \wedge \quad T - \{ r_i \} \not\models e$

A theory T is reduced for e iff

$T \models e \quad \wedge \quad \neg\exists\, r_i \in T$ such that it is not necessary for e.

S_1, S_2 are alternative models of T for e iff

$S_1 \subset T$, $S_2 \subset T$, $S_1 \neq S_2$ and S_1, S_2 are reduced for e.

We define *Model*(e, T) as the set of alternative models for example e with respect to T.

We define $Model_r(e, T)$ as the set of alternative models for example e with respect to T that contain r. Formally,

$$Model_r(e, T) = \{ S \subset Model\,(e, T) \wedge r \in S \}.$$

Reinforcement (2 of 2)

DEFINITION 3. **Pure Reinforcement.**

The pure reinforcement $\rho\rho(r)$ of a rule r from a theory T wrt. to some given observation $E = \{ e_1, e_2, \ldots, e_n \}$ is computed as the number of models of e_i where r is used.

If there are more than one model for a given e_i, *all* of them are reckoned. In the same model, a rule is computed once. Formally,

$$\rho\rho(r) = \Sigma_{i=1..n} \, \mathrm{card}(Model_r(e_i, T))$$

DEFINITION 4. **Normalised Reinforcement**

$$\rho(r) = 1 - 2^{-\rho\rho(r)}.$$

To avoid fantastic concepts:

DEFINITION 5. **Reinforcement wrt. the Data.**

The *course* $\chi(f)$ of a given fact f wrt. to a theory is computed as the product of all the reinforcements $\rho(r)$ of all the rules r used in the model of f. If a rule is used more than once, it is computed once. If f has more than one model, we select the greatest course. Formally,

$$\chi(f) = \max_{S \subset Model(f, T)} \{ \ \Pi_{r \in S} \, \rho(r) \ \}$$

Characteristics of Reinforcement

👍 Advantages:

- Reinforcement is easy to compute and allows a flexible evaluation of a theory and the data it covers.
- It provides a measure of the predictive accuracy or assumption feasibility.
- It works for evidences with noise.

👎 Drawbacks:

- For infinite evidence it must be approximated by using a finite sample.

Selection Criteria using Reinforcement

- The Most Reinforced One: The greatest *mean* ($m\chi$) of the courses of all the data presented so far.
- More Compensated: a *geometric mean* instead.

DEFINITION 6. A theory is worthy iff $m\chi(T,E) \geq 0.5$.

If the language is expressible enough there is always a worthy theory for every evidence (just choose every example as an extensional rule).

Easy to define another criteria:
- Intensional: all facts should have a course value greater than the mean divided by a constant (no exceptions).
- Consilience can be better studied: a theory is well-separable if $m\chi$ is not decreased after separation.

Examples:

EXAMPLE 2: (using equational theories)

Consider the following evidence e_1–e_{10}:

E = {	e_1: e(4) → true,	e_2: e(12) → true,
	e_3: e(3) → false,	e_4: e(2) → true,
	e_5: e(7) → false,	e_6: e(7) → false,
	e_7: e(20) → true,	e_8: e(0) → true,
	e_9: o(3) → true,	e_{10}: o(2) → false }

where natural numbers are represented as e.g. s(s(s(0))) means 3.

T_a= {	e(s(s(X)) → e(X)	: 7	0.992
	e(0) → true	: 5	0.969
	e(s(0)) → false	: 3	0.875
	o(s(s(X)) → o(X)	: 2	0.75
	o(0) → false	: 1	0.5
	o(s(0)) → true	: 1	0.5 }

The courses are $\chi(e_1, e_2, e_4, e_7, e_8)$ = 0.992 · 0.969 = 0.961, $\chi(e_3, e_5, e_6)$ = 0.992 · 0.875 = 0.868, $\chi(e_9)$ = 0.75 · 0.5 = 0.375 and $\chi(e_{10})$ = 0.75 · 0.5 = 0.375. **The mean course $m\chi$ is 0.8159.**

T_b= {	e(s(s(X)) → e(X)	: 9	0.998
	e(0) → true	: 6	0.984
	e(s(0)) → false	: 4	0.938
	o(X) → not(e(X))	: 2	0.75
	not(true) → false	: 1	0.5
	not(false) → true	: 1	0.5 }

The courses are $\chi(e_1, e_2, e_4, e_7, e_8)$ = 0.998 · 0.984 = 0.982, $\chi(e_3, e_5, e_6)$ = 0.998 · 0.938 = 0.936, $\chi(e_9)$ = 0.75 · 0.5 · 0.998 · 0.938 = 0.351 and $\chi(e_{10})$ = 0.75 · 0.5 · 0.998 · 0.984 = 0.368. **The mean course $m\chi$ is 0.8437.**

Computational Consilience

DEFINITION 7. A theory T is partitionable wrt. an evidence E iff $\exists T_1, T_2 : T_1 \subset T, T_2 \subset T$ and $T_1 \neq T_2$ such that

$$\forall e \in E : T_1 \models e \vee T_2 \models e \,.$$

We define $E_1 = \{ e \in E : T_1 \models e \}$ and $E_2 = \{ e \in E : T_2 \models e \}$ and $E_{12} = E_1 \cap E_2$.

We will use the term $S\chi(T_1 \oplus T_2, E)$ to denote:

$$m\chi(T_1, E_1)\cdot[\text{card}(E_1)-\text{card}(E_{12})/2] + m\chi(T_2, E_2)\cdot[\text{card}(E_2)-\text{card}(E_{12})/2\,]$$

DEFINITION 8. A theory T is consilient wrt. an evidence E iff there does not exist a partition T_1, T_2 such that:

$$S\chi(T_1 \oplus T_2, E) \geq m\chi(T, E) \cdot \text{card(E)}.$$

In other words, a theory T is consilient wrt. an evidence E iff there does not exist a bipartition $P \in \wp(T)$, such that every example of E is still covered separatedly without loss of reinforcement.

Computational Consilience (Example)

EXAMPLE 3:

E = { p(a), p(b), p(e), q(a), q(b), q(e), q(f) }

P = { p(X) : ρ = 0.875

q(X) : ρ = 0.9375 } $m\chi(E,P)$ = 0.9107

The partition:

P_1 = { p(X) : ρ = 0.875 } $m\chi(E_1,P_1) = 0.875$

P_2 = { q(X) : ρ = 0.9375 } $m\chi(E_2, P_2) = 0.9375$

$S\chi(P_1 \oplus P_2, E) = m\chi(E_1,P_1) \cdot 3 + m\chi(E_2, P_2) \cdot 4 = m\chi(E, P) \cdot \mathrm{card}(E) = 0.9107 \cdot 7.$

P is not consilient.

Definition 8 can be parameterised by introducing a consilience factor.

Intrinsical Exceptions and Consilience

An intrinsecal exception or extensional patch is defined as a rule r with $\rho = 0.5$, i.e. a rule that just covers one example e.

We must distinguish between:

- *completely extensional exceptions*, when r does not use any rule from the theory to cover e,
- partially extensional exceptions when r uses other rules to describe e.

> **Theorem 1.** If a worthy theory T for an evidence E has a rule r with $\rho = 0.5$, and completely extensional, then T is not consilient.

This justifies the use of consilience as the motor or maxim of theory formation:

If part of the evidence is covered extensionally, a revision should be made to conciliate it with the rest.

For explanatory induction, not only prediction errors or anomalies motivate theory revision.

Proof of Theorem 1

Just choose the partition $T_1 = T - r$ and $T_2 = T$. Since $\rho = 0.5$ then r is only used by one example e_r. Since it is a completely extensional exception, we have that r does not use any rule from T_1 to cover e_r, so $\rho'(r_i) = \rho(r_i)$ for all $r_i \in T_1$. Let n be the number of the examples of the evidence E. Hence, $m\chi(T_1, E_1) = [m\chi(T, E) \cdot n - \chi(e_r, T)\ ] / (n-1) = [m\chi(T, E) \cdot n - ½\ ] / (n-1) = [m\chi(T, E) \cdot n + m\chi(T, E) - m\chi(T, E) - ½\ ] / (n-1) = m\chi(T, E) + [m\chi(T, E) - ½] / (n-1)$.

From def. 7, the inequality simplifies as follows:

$$S\chi(T_1 \oplus T_2, E) =$$
$$m\chi(T_1, E_1) \cdot [\ \mathrm{card}(E_1) - \mathrm{card}(E_{12})/2\ ] + m\chi(T_2, E_2) \cdot [\ \mathrm{card}(E_2) - \mathrm{card}(E_{12})/2\ ] =$$
$$\{\ m\chi(T, E) + [\ m\chi(T, E) - ½] / (n-1)\ \} \cdot [\ (n-1) - (n-1)/2\ ] + m\chi(T, E) \cdot [\ n - (n-1)/2\ ] =$$
$$m\chi(T, E) \cdot [\ (n-1) - (n-1)/2 + n - (n-1)/2\ ] + [\ m\chi(T, E) - ½\ ] \cdot [\ (n-1) - (n-1)/2\ ] / (n-1) =$$
$$m\chi(T, E) \cdot [\ n\ ] + [\ m\chi(T, E) - ½\ ] / 2$$

Since T is worthy, then $m\chi(T, E) \geq 0.5$., and finally

$$S\chi(T_1 \oplus T_2, E) \geq m\chi(T, E) \cdot n = m\chi(T, E) \cdot \mathrm{card}(E).\ \square$$

Consilience and Coherence

Thagard's *modern* view of coherence is equivalent to constraint satisfaction wrt. the background knowledge *B*.

Coherence allows the analysis of:

- Deductive Compatibility with *B*.
- Explanatory Compatibility with *B*. (abduction or nomological induction)

However, it is not constructive, so it is not useful for non-nomological induction → the idea of satisfaction is not clear...

Consilience represents this notion of 'accordance' with the background knowledge *B*, by measuring the direct or constructive inter-relation with *B*.

Conclusions

Formalising 'Consilience':

- First approach to consilience based on model partition.
- Second approach based Reinforcement: Further detail on the relation between hypothesis and evidence.

From here we have shown that:

- Consilience is different (but related) that intensionality.
- Consilience can be used to detect which parts of the theory are weak.
- Consilience and Coherence are somehow *complementary*.

A consilient model is the goal of theory construction. Exceptional or unconsilient parts should trigger

Notes:

* Slide "Towards Computational Consilience"
  (If T was a Logic Theory, $M^{+}(T)$ could be the Minimal Herbrand Model of T).