A Computational Definition of 'Consilience'

José Hernández-Orallo[*]

Universitat Politècnica de València
Departament de Sistemes Informàtics i Computació,
Camí de Vera 14, Aptat. 22.012 E-46071, València, Spain
E-mail: jorallo@dsic.upv.es.

Abstract. This paper defines in a formal and computational way the notion of 'consilience', a term introduced by Whewell in 1847 for the evaluation of scientific theories. Informally, as it has been used to date, a model or theory is 'consilient' if it is predictive, explanatory and unifies the evidence. Centred in a constructive framework, where new terms can be introduced, an initial and more classical approach based on the inverse idea of separation soon manifests the problem of the introduction of *fantastic* concepts to unify disparate sub-theories. Our second approach is constructed by using a detailed evaluation of the relationship between the theory and the evidence by means of reinforcement propagation. With the use of reinforcement, fantastic concepts can be better detected and the role of consilience for theory construction and revision can be specialised for different inference mechanisms like explanatory induction, abduction, deduction and analogy.

Keywords: Consilience, Model and Theory Formation, Explanatory Induction, Abduction, Reinforcement, Analogy, Coherence, Philosophy of Science.

1. Introduction

In 1847, Whewell coined a new term, 'consilience', to comprise the relevant basics in scientific theories: prediction, explanation and unification of fields. Since all of these criteria are desirable, consilience was informally introduced as a fundamental issue for theory construction and modelling. However, a unified, formal and computational definition has not been presented

[*] Also at the Department of Logic and Philosophy of Science of the University of Valencia. On-line papers: http://www.dsic.upv.es/~jorallo/escrits/escritsa.htm

to date, integrating in a consistent way prediction, explanation and unification of fields, allowing the growth and revision of knowledge.

Throughout the paper we will deal with the process of non-deductive or hypothetical inference, i.e., the reasoning process usually represented by Science (or by everyday learning and explanation). Given some evidence E composed of facts, the goal is to obtain a theory T which explains E or/and allows the prediction of future facts. A 'bias' β is the expressive framework where hypotheses can be constructed on. The complexity of *learning* is directly related with the specificity of the bias and the background knowledge B, which is usually expressed under the same bias as the hypotheses.

Usually, we use the term *theory* to comprise the hypothesis H jointly with the necessary auxiliary concepts from the background knowledge. We will use the term *model* to designate a theory which introduces new constructed terms or extends the vocabulary of the bias. For this to happen the bias must be flexible enough to allow the creation of concepts (also known as predicate invention) and it must perform some kind of abstraction.

Our goal is precisely to define a measure of consilience for constructive languages, where new terms can be introduced or *created*. The idea of unification is straightforward when the hypothesis vocabulary is included in the vocabulary of the background knowledge and the evidence because 'fantastic' new concepts are restricted. The same does not hold, however, for constructive languages.

In the following, we will work with representational languages which are composed of rules, components, chunks or whatever other recognisable parts. Finally, we will denote that a theory T covers an example e by $T \models e$

2. Distinguishing Consilience

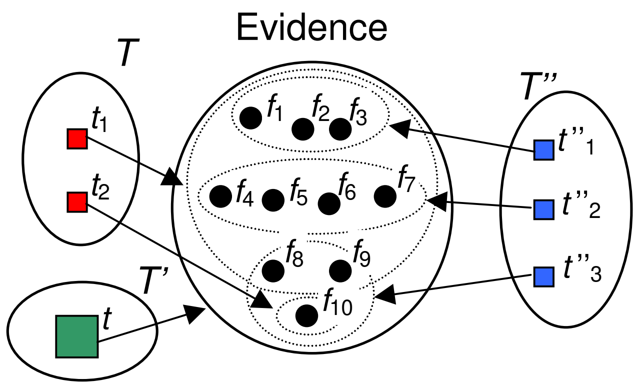
Before trying to define consilience we must distinguish it from other very related concepts.

There are evaluation criteria which are intrinsic, i.e., the theory can be evaluated by exclusively regarding to the hypothesis, like the MDL principle (or Occam's Razor formalised). However, consilience is a 'structural' criterion, because it studies how the hypothesis covers the evidence. More concretely, consilience is mainly characterised by a unified covering of the evidence or, in misspelled words, the evidence is 'consiliated' by the theory. There are different ways to evaluate this 'conciliation'. One can measure the consilience of a hypothesis H alone (more appropriate for induction) or one can measure the consilience of the $H + B$ (more appropriate for abduction).

The first trait of consilience is prediction. The predictions of a consilient theory must be plausible, so fantasies should be avoided. Moreover, it must allow the prediction of future cases, so extensional definitions should be not permitted. This has motivated some confusion between intensionality, seen as an intolerance of partial extensionality or exceptions, and consilience. The following example clarifies the difference between intensionality and consilience:

EXAMPLE 2.1.

Given the evidence E= { $f_1, f_2, \ldots f_{10}$ } and the following hypotheses: T= { t_1, t_2 }, T'= {t''} and T''= {t'_1, t'_2, t'_3}



We can observe these two different (but closely related) notions:

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

The second trait of consilience is explanation. Therefore consilience has always been alluded in the context of scientific explanation or explanatory induction (Harman 1965, Hempel 1965, Ernis 1968). Moreover, one of the important traits of abduction, seen as the inference to the best explanation, is that the abductive hypothesis (known as assumption) must be the most 'compliant' with the background knowledge. This can be identified with the notion of 'coherence' (Thagard 1978). We will discuss more deeply the relationship between coherence and consilience in section 8.2.

The third and more distinguished trait in consilience, unification, is very close to the principle of 'Common Cause' (Reichenbach 1956). However, consilience is a criterion which does not deal with causation or time dependencies, just 'uses' dependencies, and that simply prefers 'unifying' theories over separate ones.

3. Towards Computational Consilience

From a semantic point of view, a theory is characterised by the data it covers. Whatever the representational language, we will denote with $Ext(T)$ the extension, scope or covering of a theory T, i.e., $Ext(T) = \{ f : T \models f \}$. Sometimes, the terminology is confuse, because, for logical theories, $Ext(T)$ corresponds to the minimal Herbrand *model* of T.

From this elementary start point, we could investigate a purely semantic definition of consilience, based on its contrary notion, the idea of separation.

DEFINITION 3.1. Separable Theories

A theory T is n-separable in the partition of *different* theories $\Pi = \{ T_1, T_2, \dots , T_n \}$ iff $Ext(T) = \cup_{i=1..n} Ext(T_i)$ and $\forall_{i=1..n} Ext(T_i) \neq \varnothing$.

However, from this definition, we can specialise the notion of separation in many different ways, giving the following *modes* of separation:

I. *non-empty*: Exactly as DEFINITION 1.
II. *non-subset*: DEF 1 and $\forall_{i..j=1..n} (P_i \subseteq P_j \Rightarrow i=j)$.
III. *disjoint*: DEF 1 and $\forall_{i..j=1..n} (P_i \cap P_{ji} = \varnothing)$.
IV. *non-subset <u>extension</u>*: DEF. 1 and $\forall_{i..j=1..n} (Ext(P_i) \subseteq Ext(P_j) \Rightarrow i=j)$.
V. *disjoint <u>extension</u>*: DEF 1 and $\forall_{i..j=1..n} (Ext(P_i) \cap Ext(P_j) = \varnothing)$.

If we define a theory as consilient iff it is not separable, the preceding five modes give five characterisations of consilient theories.

EXAMPLE 3.2. (Using Horn Theories):

P_1= { p(a). q(X) :- r(X). r(a). } is separable for all modes into Π = { { p(a) } , { q(X) :- r(X). r(a) }}.

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

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

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

P_5= { s(X) :- p(X), q(b). p(X) :- q(X). t(X) :- p(X), q(a) } is non-subset (extension) and disjoint extension separable 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.

Several problems can be detected from the previous example. Modes I, II, and IV are so strict that they do not allow any modularity at all. On the contrary, modes III, IV, and V can be 'conciliated' in a tricky way. In example

3.2, for instance, 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 modes iii-iv.

In our opinion, a strictly semantic approach is not sufficient for defining consilience. Accordingly, the next section presents a structural approach.

4. Reinforcement

For our goal of defining consilience, it is more appropriate to establish in further detail the relation between the hypothesis and the evidence. Furthermore, it would be more accurate to talk about a degree of consilience instead of 'consilient' or 'unconsilient' theories.

In (Hernandez-Orallo 1999) several theory analysis and evaluation measurements are presented based on the idea of reinforcement. The idea of reinforcement to validate a theory has been supported by many psychological studies on ontology and epistemology. Whatever the approach to knowledge construction, the construction or revision of knowledge must come from a gain or loss, respectively, of reinforcement, also known as apportionment of credit (Holland et al. 1986). From (Hernandez-Orallo 1999), we adapt in this section the basic constructions to compute the reinforcement degree for a given theory, depending on past observations, and the evidence from the point of view of the evidence itself.

DEFINITION 4.1. Necessary Component

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

DEFINITION 4.2. Reduced Theory

A theory T is reduced for e iff $T \models e \;\wedge \neg\exists\, r_i \in T$ such that it is not necessary for e.

We will say that two sub-theories 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. From here, we can define $Models(e, T)$ as the set of alternative models for example e with respect to T.

DEFINITION 4.3. Alternative Models

$$Models(e, T) = \{ S \subset T : S \text{ is reduced for } e \}.$$

We can particularise definition 4.3. by defining $Models_r(e, T)$ as the set of alternative models for example e with respect to T that contain r. Formally,

DEFINITION 4.4. Alternative Models which contain r:

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

With these definitions, it is straightforward to define reinforcement.

DEFINITION 4.5. 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,

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

DEFINITION 4.6. Normalised Reinforcement

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

The last definition is motivated by the convenience of a measure between 0 and 1.

From these definitions some properties are proven in (Hernandez-Orallo 1999). For instance, the most reinforced theory is not the shortest one in general, but, *in the limit*, simplicity is a good criterion to obtain consilience. However it is important to remark that, somehow surprisingly, even some kind of redundancy (*investment*) does not necessarily imply a loss of reinforcement ratio.

Nonetheless, this measure of reinforcement of the *theory* could unfairly increase by the introduction of *fantastic* concepts. The rationale relies on the fact that an invented rule, used in every other rule of the theory, could *unjustifiably* increase the reinforcement ratio of a theory. Since it is difficult to detect whether these rules are invented or not, simplicity is a reasonable criterion to avoid these fantastic concepts. However, there is a different way out to measure the validation *wrt. the data*:

DEFINITION 4.7. Reinforcement wrt. the Data.

The *course* $\chi(f)$ of a given fact f wrt. 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,

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

With this definition, it is proven in (Hernandez-Orallo 1999) that no fantastic rule can be added in the previous way, but the good properties of the original definition are still preserved.

5. Selection Criteria

Once the grounds of the theory are ensured by the measurement of the course of the evidence instead of the normalised reinforcement of the theory, we can construct different selection criteria. The first idea is to select the theory T with the greatest *mean* of the courses of all the data (evidence E)

presented so far, denoted by $m\chi(T,E)$. If the language is expressible enough there is always a theory for every evidence (just choose every example as an extensional rule) and this extensional theory has $m\chi(T,E) = 0.5$. In the following we will say that a theory is *worthy* iff $m\chi(T,E) \geq 0.5$.

In explanatory induction, however, it is not sufficient to force a great mean. In order to obtain a more compensated theory, a *geometric mean* can be used instead,. Even more, any anomaly should be banned. Consequently, one would discard theories where a fact has a course value less than the mean divided by an intensionality constant.

The following example shows the use of $m\chi$. However, other of the criteria which have been commented could also be used.

EXAMPLE 5.1. (using equational theories)

Consider the following evidence e1–e10:

$E = \{$	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.

$T_a = \{$	r_{a1}: e(s(s(X)) → e(X)	: 7	0.992
	r_{a2}: e(0) → true	: 5	0.969
	r_{a3}: e(s(0)) → false	: 3	0.875
	r_{a4}: o(s(s(X)) → o(X)	: 2	0.75
	r_{a5}: o(0) → false	: 1	0.5
	r_{a6}: o(s(0)) → true	: 1	0.5 $\}$

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

$T_b = \{$	r_{b1}: e(s(s(X)) → e(X)	: 9	0.998
	r_{b2}: e(0) → true	: 6	0.984
	r_{b3}: e(s(0)) → false	: 4	0.938
	r_{b4}: o(X) → not(e(X))	: 2	0.75
	r_{b5}: not(true) → false	: 1	0.5
	r_{b6}: not(false) → true	: 1	0.5 $\}$

The courses are $\chi(e_1, e_2, e_4, e_7, e_8) = 0.998 \cdot 0.984 = 0.982$, $\chi(e_3, e_5, e_6) = 0.998 \cdot 0.938 = 0.936$, $\chi(e_9) = 0.75 \cdot 0.5 \cdot 0.998 \cdot 0.938 = 0.351$ and

$\chi(e_{10}) = 0.75 \cdot 0.5 \cdot 0.998 \cdot 0.984 = 0.368$. The mean course $m\chi$ is 0.8437.

This example provides more insight in our goal of defining consilience. T_a can be split without loss of reinforcement because there are no shared rules between the definition of *odd* and the definition of *even*. However T_b has been more 'conciliated' by the use of a new invented term (in this case negation), which makes that it cannot be separated without loss of reinforcement. The following section formalises this idea.

6. Computational Consilience

The idea of separation is still necessary for any definition of consilience:

DEFINITION 6.1.

A theory T is divisible 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 $\forall e \in E : T_1 \models e \vee T_2 \models e$.

However, it is not sufficient, as we have seen. We will use the following notation $E_1 = \{ e \in E : T_1 \models e \}$, $E_2 = \{ e \in E : T_2 \models e \}$, $E_{12} = E_1 \cap E_2$, and finally we will use the term $S\chi(T_1 \oplus T_2, E)$ to denote $m\chi(T_1, E_1) \cdot [card(E_1) - ½ \cdot card(E_{12})] + m\chi(T_2, E_2) \cdot [card(E_2) - ½ \cdot card(E_{12})]$.

DEFINITION 6.2.

A theory T is consilient wrt. an evidence E iff there does not exist a partition T_1, T_2 such that: $S\chi(T_1 \oplus T_2, E) \geq m\chi(T, E) \cdot card(E)$.

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

For example 5.1, T_a is divisible into $T_{1a} = \{ r_{a1}, r_{a2}, r_{a3} \}$ and $T_{2a} = \{ r_{a4}, r_{a5}, r_{a6} \}$ and $S\chi(T_{1a} \oplus T_{2a}, E) = 0.9261 \cdot [8 - ½ \cdot 0] + 0.375 \cdot [2 - ½ \cdot 0] = 8.159 = m\chi(T_a, E) \cdot 10$. In this way, T_a is not consilient. On the contrary, it can be shown that there is no partition of T_b to make true the disequality of definition 6.2.

The next example shows that consilience is again a delicate notion:

EXAMPLE 6.1. (using Horn theories)

Consider the following extensional theory T= { p. q. } for the following simple theory E= { p, q }. As expected, $m\chi(T, E) = (0.5 + 0.5) / 2 = 0.5$ and by using the partition T_1= { p. }, T_2= { q. } is easy to show that it is not consilient.

The trick is again the addition of a new fantastic rule *f* in the following way: T'= { p:- f. q:- f. f }. As we have said, the mean course is robust to this kind of tricks, and it is clearly lower: $m\chi(T', E) = (0.5 \cdot 0.75 +$

$0.5 \cdot 0.75) / 2 = 0.375$. However, the only partition which is now possible, T'_1= { p:- f. f }, T'_2= { q:-f. f. } gives that $S\chi(T'_1 \oplus T'_2, E) = 0.25 \cdot [1 - ½ \cdot 0] + 0.25 \cdot [1 - ½ \cdot 0] = 0.5 < m\chi(T', E) \cdot 2$. The result is that T' is consilient!

This example can be interpreted in two ways. If one has T and tries to make it consilient by using a fantastic concept, she would get an important decrease in $m\chi(T', E)$ enough for discarding T'. On the other hand, if one considers T' from scratch (without knowing T), she could be cheated by the illusion that T' is a good consilient theory if these invented concepts were difficult to detect.

It is important to realise that definition 6.2. is *reliable*; independently from whether the unifying concept would be fantastic or not, the theory is properly consilient.

The aftermath harmonises with the classical rationale of the plausibility of a theory: it depends on the intuition, intelligence or whatever other ability to unveil fantasies by comparing the current theory with other competing theories. The advantage of our measures of mean course and consilience based on reinforcement is that the first one avoids fantastic concepts, so giving an approximation to plausibility, which must be weighed up with consilience.

The following example shows the use of $m\chi$ and consilience in the context of abduction and background knowledge. In this case, invented concepts are more difficult to introduce if the background knowledge cannot be modified by adding a fantastic rule.

EXAMPLE 6.2. (using Horn theories)

Let us suppose a "linguistically debatable" background knowledge B, although we assume that it is completely validated, i.e. $\forall r \in B\ \rho(r) = 1$.

B= { r_{b1}: European(X) :- Spanish(X)
r_{b2}: European(X) :- Portuguese(X)
r_{b3}: European(X) :- French(X)
r_{b4}: European(X) :- Belgian(X)
r_{b5}: French-speaking(X) :- French(X)
r_{b6}: French-speaking(X) :- Walloon(X), Belgian(X)
r_{b7}: English-speaking(X) :- British(X)
r_{b8}: Portuguese-speaking(X) :- Portuguese(X)
r_{b9}: Spanish-speaking(X) :- Spanish(X)
r_{b10}: Spanish-speaking(X) :- Mexican(X) }

Given the following evidences, some hypotheses can be abduced:

E_1 = { e_1: European(Mauro), e_2: Portuguese-speaking(Mauro) }

$h_1 = E_1$ $m\chi(B+h_1, E_1) = 0.5$

$h_2 = \{$ Portuguese(Mauro). $\}$ $m\chi(B+h_2, E_1) = 0.75$

Moreover h_2 is consilient wrt. E_1.

$E_2 = \{$ e_3: European(Luis), e_4: Spanish-speaking(Luis) $\}$

$h_3 = E_2$ $m\chi(B+h_3, E_2) = 0.5$

$h_4 = \{$ Spanish(Luis). $\}$ $m\chi(B+h_4, E_2) = 0.75$

$h_5 = \{$ Mexican(Luis). European(Luis). $\}$ $m\chi(B+h_5, E_2) = 0.5$

Only h_4 is consilient

$E_3 = \{$ e_5: European(Daniel), e_6: French-speaking(Daniel) $\}$

$h_6 = E_4$ $m\chi(B+h_6, E_3) = 0.5$

$h_7 = \{$ French(Daniel). $\}$ $m\chi(B+h_7, E_3) = 0.75$

$h_8 = \{$ Walloon(Daniel). Belgian(Daniel). $\}$ $m\chi(B+h_8, E_3) = 0.625$

Both h_7 and h_8 are consilient.

$E_4 = \{$ e_6: European(Steve), e_6: English-speaking(Steve) $\}$

$h_9 = E_4$ $m\chi(B+h_9, E_4) = 0.5$

$h_{10} = \{$ British(Sue). European(Sue). $\}$ $m\chi(B+h_{10}, E_4) = 0.5$

$h_{11} = \{$ British(Sue). European(X):-British(X). $\}$ $m\chi(B+h_{11},E_4) = 0.625$

In this last case, only h_{11} is consilient, and it shows that an extension can to B can be made with new rules in order to cover the evidence in a consilient way.

However, the example shows that in many cases $m\chi$ is positively related to consilience, so it is a good criterion to guide knowledge creation and revision. Abduction has been naturally incorporated as a special case of explanatory induction, where, in general, the hypotheses are *factual* (although in the examples h_{11} includes non-factual ones and it can also be considered an abduction).

Finally, definition 6.2 can be parameterised by introducing a consilience factor:

DEFINITION 6.3.

The degree of consilience of a theory T wrt. an evidence E is defined as the minimum real number k such that there exists a partition T_1, T_2 such that: $k \cdot S\chi(T_1 \oplus T_2, E) \geq m\chi(T, E) \cdot card(E)$.

From the computational point of view, both $m\chi$ and consilience degree should be computed jointly, in order to reduce the number of partitions which are to be examined.

7. Inference Processes

Explanatory induction has been distinguished as the major process to obtain consilient theories. Consilience can play an important role in induction under this schema:

Explanatory Induction:

Background Knowledge: empty or used auxiliarily.

Evidence: E_1 and E_2.

Process: Construct a unified theory A for E_1 and E_2.

Where A should comply with the restrictions of consilience and plausibility ($m\chi$).

Similarly, as we have seen in the examples of the previous section, abduction fits naturally by a more important use of the background knowledge:

Abduction:

Background Knowledge: a fact b entails E_1 and E_2.

Evidence: E_1 and E_2.

Process: Assume b to ensure consilience.

Although induction and abduction are recognised as the basic processes in scientific discovery, there is an inference process which is the fundamental mechanism for obtaining consilient theories, analogy. The reason is simple: analogy extracts a common superstructure between two situations, and this 'shared' superstructure is reinforced by both situations.

Analogy:

Background Knowledge: b entails E_1 and c entails E_2.

Evidence: E_1 and E_2.

Process: Extract similarities between b and c into a new superstructure a in order to obtain a consilient theory composed of a, b' and c'.

We can state that analogy favours consilience.

Theorem 7.1. If b entails E_1, c entails E_2, b does not entail E_1 and c does not entail E_2, the new theory $T = \{ b', c', a \}$ such that $T_1 = \{ b', a \} \models E_1$ and $T_2 = \{c', a \} \models E_2$, and no other proper subset of T covers any example, then T is consilient.

Proof. Since no other proper subset of T covers any example but T_1 and T_2, then there is only one possible partition to study consilience $\{T_1, T_2\}$. Since E_1 and E_2 are non-empty, then $m\chi(a, E_1) < m\chi(a, E_1 \cup E_2) > m\chi(a, E_2)$, and then $S\chi(T_1 \oplus T_2, E) < m\chi(T, E_1 \cup E_2) \cdot card(E_1 \cup E_2)$. From definition 6.2, T is consilient. □

Once again, analogy, as it has defined, allows the introduction of fantastic concepts. In order to talk about a 'real' analogy, some information must be shared between b and c and moved into a. In other words, b' and c' should be simplified wrt. b and c. This can be related with reinforcement and extended from simple components like b and c to sub-theories composed of many rules or components.

DEFINITION 7.2. Non-fictitious Analogy

Consider a theory T covering E, i.e., $\forall e \in E, T \models e$, which contains two sub-theories T_1 and T_2, which cover $E_1 \subset E$ and $E_2 \subset E$, respectively. A non-fictitious analogy is the addition to T of a new super-theory A, and the modification of T_1 and T_2 into T'_1 and T'_2 such that $T' = ((T / T_1) / T_2) \cup A \cup T'_1 \cup T'_2$ covers E, i.e. $\forall e \in E, T' \models e$, with the additional conditions that $m\chi(T', E) \geq m\chi(T, E)$ and T' must be consilient wrt. E_1 and E_2.

This definition is more accordant with classical computational approaches to analogy (Kling 1971, Winston 1992).

Finally, there is another process which is important for obtaining consilience. If the theory is not omniscient, i.e., everything that can be ever deduced is effectively deduced by the system, we have that deduction can be also a source of consilience.

Non-omniscient Deduction:

Background Knowledge: b entails E_1, c entails E_2 and a. No relation is still established among b, c and a.

Evidence: E_1 and E_2.

Process: Show that a entails both b and c.

It is important to highlight the difference between non-omniscient deduction and computation (or deterministic proof systems). The first one can be informative and creative and it can connect two unrelated things, so increasing reinforcement, and, in many cases, it can unify separate theories. In this way, induction and abduction should not be seen as inverse processes of deduction, *in terms of information gain*. The deductive-nomological model of explanatory induction introduced in 1949 by Hempel and Oppenheim (Hempel 1965), can be better seen as a computational-nomological one, because any inference process like induction, deduction, abduction and analogy takes place in a computational system.

8. Related Concepts

In the beginning we have commented on some related concepts to consilience, especially intensionality and coherence. In this section we study in further detail the differences and similarities, once consilience has been more concretely defined.

8.1. Intrinsic Exceptions and Consilience

It is easy to define an intrinsic exception or extensional patch as a rule r with $\rho = 0.5$, i.e. a rule that just covers one example e. Nonetheless, 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.

It is possible to establish clearly the relation between the former and consilience. The latter, however, are also usually conflicting to consilience.

Theorem 8.1. 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.

Proof. 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 definition 6.2, 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 $S\chi(T_1 \oplus T_2, E) \geq m\chi(T, E) \cdot n = m\chi(T, E) \cdot \mathrm{card(E)}$. □

This theorem justifies the avoidance of exceptions in order to obtain consilient theories. In the process of theory construction, if a new evidence is covered extensionally, the theory necessary loses its consilience and revision must be done in order to 'conciliate' this new evidence with the previous theory. This means that, for explanatory induction, not only prediction errors or anomalies (consistency) but consilience can trigger theory revision.

8.2. Consilience and Coherence

Coherence has been advocated as the key issue in scientific explanation (Thagard 1978) and abduction (Ng. and Mooney, 1990). An explanation is coherent with the evidence and the background knowledge if it is the most compatible, in the way that it confirms more positive items from the background knowledge and the evidence, and it activates less negative items. Recently, this idea has been identified with constraint satisfaction (see e.g. Thagard and Verbeurgt, 1997 or Thagard 1998), although the term has been generally used in a broader sense (Thagard 1989).

Despite their close relationship, we think that our definition of consilience has two differentiated issues wrt. coherence. First, the idea of unification is not *explicitly* present in coherence, and, secondly, the idea of coherence as constraint satisfaction is difficult to extend to constructive languages, because new introduced or invented terms are difficult to identify with positive or negative constraints (in some way, the vocabulary is extended with new constraints).

In our opinion, we think that our measurement of mean course $m\chi$ matches this notion of 'constructive' coherence. To be more precise, however, our definition should be extended with negative reinforcement. Consequently, coherence and consilience would be connected in the same way as we discussed that mean course and consilience were connected.

9. Conclusions

In this paper we have addressed formally and computationally the notion of consilience for constructive languages. Pure semantic approaches based on model partition present many problems of introduction of fantastic concepts. A second approach based on reinforcement allows further detail on the relation between hypothesis and evidence, and these fantastic concepts are much more easier to detect.

Different inference processes have been re-understood in the context of consilient theory construction. Explanatory induction, abduction, analogy and even deduction are valuable tools for obtaining consilient theories.

The most important result is that consilience has been related to and differentiated from many other classical notions in explanatory induction and scientific discovery, like avoidance of anomalies and coherence. Moreover, it has been shown that, under consilience considerations, theory revision should also be triggered by unconsilient parts and not only by inconsistencies or anomalies.

Acknowledgements

I am grateful to Paul Thagard for clarifying a misinterpretation of his theory of coherence. Section 8.2 has been remade accordingly.

References

Barker, S.F., 1957, *Induction and Hypothesis*, Ithaca.

Bar-Hillel, Y. and Carnap, R., 1953, Semantic information, *British J. for the Philosophy of Science* 4:147-157.

Botilier, C. and Becher, V., 1995, Abduction as belief revision, *Artificial Intelligence* 77:43-94.

Carnap, R. 1952, *The Continuum of Inductive Methods*, University of Chicago, Chicago, IL.

Dietterich, T.G. and Flann, N.S., 1997, Explanation-based learning and reinforcement learning: a unified view, *Machine Learning*, 28:169-210.

Ernis, R., 1968, Enumerative induction and best explanation, *J. of Philosophy*, LXV (18): 523-529.

Flach, P. 1996, Abduction and induction: syllogistic and inferential perspectives, in: Working Notes of the ECAI'96 Workshop on *Abductive and Inductive Reasoning*, M. Denecker, L. De Raedt, P. Flach and T. Kakas, eds. pp. 7-9, Brighton.

Flach, P. and Kakas, A. (eds), 1998, *Abduction and Induction. Essays on their Relation and Integration*, in press, Kluwer.

Harman, G., 1965, The inference to the best explanation, *Philosophical Review*, 74: 88-95.

Hempel, C.G., 1965, *Aspects of Scientific Explanation*, The Free Press, New York.

Hendricks, V.F. and Faye, J., 1998, Abducting explanation, extended abstract in: MBR'98 Abstracts, S. Rini and G. Poletti, eds., complete paper, Department of Philosophy, Univ. of Copenhagen.

Hernandez-Orallo, J., 1999, Constructive reinforcement learning, *International Journal of Intelligent Systems*, to appear.

Hernandez-Orallo, J. and Garcia-Varea, I., 1998, Distinguishing abduction and induction under intensional complexity, in: Proceedings of the ECAI'98 Workshop on Abduction and Induction in AI, Flach, P.; Kakas, A., eds., 41-48, Brighton.

Hernandez-Orallo, J. and Ramirez-Quintana, M.J., 1998a, Inductive inference of functional logic programs, in: CompulogNet Area Meeting on Computational

Logic and Machine Learning, J. Lloyd, ed., at the 1998 Joint International Conference and Symposium on Logic Programming (JICSLP'98), 49-55.

Hernandez-Orallo, J. and Ramirez-Quintana, M.J., 1998b, Inverse narrowing for IFLP, in: Proceedings of the 1998 Joint Conference of Declarative Programming, APPIA-GULP-PRODE'98, Freire-Nistal, J.L.; Falaschi, M.; Vilares-Ferro, M., eds., 379-392.

Hintikka, J., 1970, Surface information and depth information, in: *Information and Inference*, Hintikka, J.; Suppes, P., eds., D. Reidel Publishing Company, 263-297.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., and Thagard, P.R., 1986, *Induction. Processes of Inference, Learning and Discovery*, The MIT Press.

Kaelbling, L., Littman, M., and Moore, A., 1996, Reinforcement learning: a survey, *J. of Artificial Intelligence Research*, 4: 237-285.

Karmiloff-Smith, A., 1992, *Beyond Modularity: A Developmental Perspective on Cognitive Science*, The MIT Press.

Kling, R.E., 1971, A paradigm for reasoning by analogy, *Artificial Intelligence*, 2:147-178.

Kuhn, T.S., 1970, *The Structure of Scientific Revolutions*, University of Chigago.

Muggleton, S. and De Raedt L., 1994, Inductive logic programming — theory and methods, J. of Logic Prog., 19-20:629-679.

Ng. H. and Mooney, R., 1990, On the role of coherence in abductive explanation, in: *Proceedings of the Eighth National Conference on Artificial Intelligence*, 337-342 Boston, MA. AAAI Press.

Peirce, C.S., 1867/1960, *Collected Papers of Charles Sanders Peirce*, Cambridge. Harvard University Press.

Popper, K.R., 1962, *Conjectures and Refutations: The Growth of Scientific Knowledge*, Basic Books, New York.

Sharger, J. and Langley, P., 1990, *Computational Models of Scientific Discovery and Theory Formation*, Morgan Kaufmman, 1990.

Thagard, P., 1978, The best explanation: criteria for theory choice, *J. of Philosophy*, 75:76-92.

Thagard, P., 1989, Explanatory coherence, *The Behavioural and Brain Sciences*, 12 (3): 435-502.

Thagard, P., 1992, Conceptual Revolutions, Princeton Univ. Pr, Princeton, N.Y.

Thagard, P., 1998, Probabilistic networks and explanatory coherence, in: Automated Abduction: Inference to the Best Explanation, P. O'Rorke and J. Josephson, eds., Menlo Park, AAAI Press.

Thagard, P. and Verbeurgt, K., 1997, Coherence as Constraint Satisfaction, Cognitive Science, forthcoming.

Whewell W., 1847, *The Philosophy of the Inductive Sciences*, New York: Johnson Reprint Corp.

Winston P.H., 1992, *Artificial Intelligence*, 3rd Edition, Addison-Wesley.