Coherence of Laws*

Rex Kwok and Norman Y. Foo
Knowledge Systems Group
Department of Artificial Intelligence
School of Computer Science and Engineering
University of New South Wales, NSW 2052
Australia

Abhaya C. Nayak
Knowledge Systems Group
School of MPCE
Macquarie University
NSW 2109
Australia

Abstract

The core of scientific theories are *laws.* These laws often make use of theoretical terms, linguistic entities which do not directly refer to observables. There is therefore no direct way of determining which theoretical assertions are true. This suggests that multiple theories may exist which are incompatible with each other but compatible with all possible observations. Since such theories make the same empirical claims, empirical tests cannot be used to differentiate or rank such theories. One property that has been suggested for evaluating rival theories is *coherence.* This was only understood qualitatively until we [Kwok, et.al. 98] introduced a coherence measure based on the average use of formulas in support sets for observations. The idea was to identify highly coherent theories with those whose formulas that are tightly coupled to account for observations, while low coherence theories contain many disjointed and isolated statements. Our current approach generalizes that insight to accommodate fundamental ideas from the philosophy of science and better mirrors scientific practice. Moreover, this new approach is neutral with respect to the philosophy and practice of science, and is able to explain notions like modularization using coherence.

1 Introduction

This extended summary highlights the main points of the paper, a full version [Kwok, et.al. 03] of which can be obtained electronically. This section motivates the problem and subsequent sections outline the definitions which formalize the intuitions behind coherence, describe some properties that flow from these definitions, and provide examples of their use.

Scientific theories evidently comprise laws that use vocabularies that contain terms which on the one hand refer to observations, and on the other refer to postulated or theoretical entities that are not directly observable. It is in fact quite common for two theories *T1* and T2 that agree on the status of their observational terms to differ in their theoretical terms. One way to compare TI and T2 is to say that TI is more coherent than T2 if in accounting for the observations the formulas in TI "work together better" than those in T2, or are "more useful" than in T2. A persuasive advocate for such properties is Bonjour iBonjour 85]. In a previous paper [Kwok, et.al. 98] we proposed a definition that amounted to a *quantitative* measure of coherence of theories. In this paper we elaborate on the definition, repairing its deficiencies and extending its range of application.

*Supported in part by the Australian Research Council.

2 Coherence

Definition 1 (Supports for Observations) *Given an input set I and and output set O, a subset TofT is a* I-relative support *for a set* O *of observations if*

1. Γ *accounts for* O, *i.e.*, $\Gamma \cup I \models \alpha$ *for each* $\alpha \in O$;
2. Γ *is minimal, i.e, for no* $\Delta \subset \Gamma$ *does* Δ *account for* O.

Let S(T, I, O) denote the set of all I-relative supports for O.

This definition differs from that in [Kwok, et.al. 98] in the relativisation of notion of support to the input set 7, which better models scientific practice.

Assumption 1 (Clausal Basis Assumption) *All bases of theories are* clauses.

Lemma 1 (Some Properties of S(T,I,O)) *Fix observation sets I and O, and let* $T_1 \subseteq T_2$ *be theories.*

1. **Monotonicity in T:** $S(T_1, I, O) \subseteq S(T_2, I, O)$
2. $T_1 \cup I \models O$ *iff* $\exists\, \Gamma \subseteq T_1$ *such that* $\Gamma \in S(T_2, I, O)$.

Fix observation set I *and theory* T. *Let* $O_1 \subseteq O_2$ *be observation sets. Then for every* $\Gamma_2 \in S(T, I, O_2)$ *there is a* $\Gamma_1 \in S(T, I, O_1)$ *such that* $\Gamma_1 \subseteq \Gamma_2$.
Fix observation set O *and theory* T. *Let* $I_1 \subseteq I_2$ *be observation sets. Then for every* $\Gamma_1 \in S(T, I, O_1)$ *there is a* $\Gamma_2 \in S(T, I, O_2)$ *such that* $\Gamma_2 \subseteq \Gamma_1$.

Definition 2 (Utility of a Formula) *The* Utility *of a formula* α, *if* $S(T, I, O) \neq \emptyset$, *in a theory* T *with respect to an I-relative observation set* O *is:*

$$U(\alpha, T, I, O) = \frac{|\{\Gamma : \alpha \in \Gamma \text{ and } \Gamma \in S(T, I, O)\}|}{|S(T, I, O)|}$$

Informally, this is the *relative frequency* of occurence of *a* in the support sets for *O.*

Definition 3 (Coherence of a Theory) *Let T be a finite theory* $\{\alpha_1, \ldots, \alpha_n\}$, *I a finite sequence of (input) observations* $(I_1, I_2, \ldots I_m)$ *and O a finite sequence of (output) observations* $(O_1, \ldots, O_m)$. *The I-relative* coherence *of T with respect to O is:*

$$C(T, I, O) = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} U(\alpha_i, T, I_j, O_j)$$

Informally, coherence is the *average* utility of the elements of T in supporting some observations with the help of others. The inputs do not figure directly in the counting because it is the *internal* laws (or rules) of *T* that we are assessing for how the outputs are supported.

3 A Typical Application

In the full paper there are applications of the above definitions to (i) demolish Craig's Trick [Craig 53], alleged by some to show that theoretical terms are unnecessary, by demonstrating that it yields highly incohrent theories; (ii) argue that Mendel's assumption of two independent theoretical characteristics to account for the observations of his pea plant experiments yields a highly coherent theory; and (iii) indicate that programs that realize the Kolmogorov complexity of sequences are maximally coherent. Here we focus on a typical application and its implication for coherence of modular theories.

3.1 The Black Swan Fix

Prior to western ornithologists exploration of Australia all the swans they had hitherto encountered were white in color. For this focussed domain, there is only one type of object, namely swans, that are of interest. The observational predicates are *swan* and *white,* and we regard the former as the input and the latter as the output. A succinct way to capture induction is the rule 1 in the theory *T* below:

$$\forall x \; swan(x) \rightarrow white(x). \qquad (1)$$

Notice that *T* does not have any theoretical terms as we have specified that both the predicates are observational. In Australia they saw black swans. Here is an ad hoc way to revise *T* minimally if we can enumerate these black swans as additions to the original input set, i.e. these new swans are *sw*, sw_2,..., sw_n.* Call this *fix* T_n. The revised rules that replace rule 1 are:
(1 sentence)

$$\forall x \; [swan(x) \wedge \bigwedge_{i \leq n} x \neq sw_i] \rightarrow white(x)] \qquad (2)$$

and *{2n* sentences)

$$\bigcup_{i \leq n} \{swan(sw_i), black(sw_i)\} \qquad (3)$$

Suppose the new observation terms are about swan *color,* i.e., black or white.

T_n has $2n+1$ sentences. For any finite set of *k* black swans, there are exactly *2k* sentences in T_n that support their color.

Each such support sentence has utility 1 for a particular swan *swi,* and 0 for other swans.

Hence the coherence of T_n for such *k* observations is $\frac{2k}{2n+1}$, which is asymptotically 0 with large *n.* This is an argument against the fix. The "good" *fix* is what happens in inductive learning when a predicate is invented to summarize the discovery that black swans live in Australia, viz., the new theory X" wit 2 sentences

$$\forall x \; [swan(x) \wedge \neg Australian(x) \rightarrow white(x)] \qquad (4)$$

$$\forall x \; [swan(x) \wedge Australian(x) \rightarrow black(x)] \qquad (5)$$

The input set I now comprises pairs of the *swan* atoms and the new observable *Australian* literals. The output *O* are the two color terms *white* and *black.* Now for any one swan (call it *c)* observation, its color is supported either by the formulas 4 and $\neg Australian(c)$, or by the formulas 5 and *Australian(c).* Therefore, irrespective of the color the support set for each observation has cardinality 2. Suppose there are *k* white and k_2 black swans in an observation sequence. It is then easy to see that the utilities of formula 4 is 1 for *k1* observations but 0 for the k_2 observations; formula 5 is the dual of the preceding. The coherence of this theory for any $k_1 + k_2$ swans is therefore $\frac{1}{2}$.

3.2 Modularization and Coherence

In the above example, suppose we partition the output observational terms into *black* (swans) and *white* (swans), denoting the disjoint sets by O_b and O_w respectively. Likewise, we partition the input set into two, I_b and I_w denoting the pairs of hypothesized Australian literal and swan atom. Then it is not hard to see that the formula 4 is in all support sets of $S(T', I_w, O_w)$, but is not in any support set of $S(T', I_b, O_b)$. Dually, the formula 5 is in all support sets of $S(T', I_b, O_b)$ but in none of those of $S(T', I_w, O_w)$. The utility of each formula is 1 in their respective support sets, and 0 in the other. This is about as strong as we can get in *modularizing* a theory.

This idea has the following obvious generalization. Suppose an observation set to be accounted for can be partitioned into $\{O_1, \ldots, O_n\}$ and the theory *T* has invented theoretical terms $\{\gamma_1, \ldots, \gamma_n\}$, such that for each i, γ_i is in every set of $S(\Gamma, I, O_i)$. Then the γ_i *modularize* the theory *T* with respect to the observation partitions.

References

[Bonjour 85] L. Bonjour. *The Structure of Empirical Knowledge.* Harvard University Press, 1985.

[Craig 53] W. Craig. On axiomatizability within a system. In *The Journal of Symbolic Logic, 18,* pages 30-32, 1953.

[Kwok, et.al. 98] R.B.H. Kwok, A.C. Nayak, N. Foo. Coherence Measure Based on Average Use of Formulas. *Proceedings of the Fifth Pacific Rim Conference on Artificial Intelligence,* 553-564, LNCS v. 1531, Springer Verlag, 1998.

[Kwok,et.al.03] R.B.H. Kwok, N. Foo and A.C. Nayak, Coherence of Theories (full version). Available as Postscipt file via FTP ftp://ftp.cse.unsw.edu.au/pub/users/ksg/Confercnce/ijcai03-Coherence.ps.gz.