Coherence of Laws

Rex Kwok and Norman Foo

The School of Computer Science and Engineering
University of New South Wales, Sydney NSW 2052, Australia
email: rkwok@cse.unsw.edu.au, norman@cse.unsw.edu.au

Abhaya Nayak

Department of Computing
Division of Information and Communication Sciences
Macquarie University, NSW 2109, Australia
email: abhaya@ics.mq.edu.au

Technical Report
UNSW-CSE-TR-0719
October 2007

THE UNIVERSITY OF
NEW SOUTH WALES

School of Computer Science and Engineering
The University of New South Wales
Sydney 2052, 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 one another 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 investigated qualitatively until we [Kwok, et.al. 98] introduced a coherence measure based on the average use of formulas in support sets for observations. Our 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. The present paper generalizes it to accommodate fundamental intuitions 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

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 entities that are not directly observable. This is the case for both the traditional "hard" sciences like physics and biology, and the modern "soft" sciences like economics and sociology. The first category of terms are *observational*, and the second category are *theoretical*. In genetics *DNA* is observational while *gene* is theoretical; in thermodynamics, *temperature* is observational but *entropy* is theoretical. In economics *interest rate* is observational while *risk* is theoretical; and in psychology *IQ score* is observational and *intelligence* is theoretical. However, philosophers [Sellars 88, van Fraassen 80] have strong misgivings about sharp distinctions between theoretical and observational terms. Indeed, among their many reasons for this is the one that has some historical justification. A number of entities that were once considered theoretical became observational with the advance of instrumentation, e.g. electron microscopy and radio telescopy. Nevertheless, it is pragmatically useful to distinguish these two categories as a way to compare alternative theories, albeit relative to a particular phase of scientific development. The formalism we propose here is flexible enough to permit arbitrary division of scientific terms into theoretical and observational components, and therefore can accommodate at least part of the philosophers' concerns while respecting the practical choices of working scientists and engineers. It can also be used to classify the different emphases of deduction, prediction, and abduction.

One justification for theoretical terms in scientific laws is that such terms provide increased *coherence* for the theories. While coherence is interpreted diversely by different researchers, there is some consensus about how to compare two theories for their degree of coherence. Here are some of the desired informal properties of coherence. If $T1$ and $T2$ are two empirically equivalent theories[1], $T1$ is more coherent than $T2$ if, in accounting for the observations, (i) the formulas in $T1$ "work together better" than those in $T2$, or (ii) the formulas in $T1$ "couple tighter" than in $T2$, or (iii) the formulas in $T1$ are "more useful" than in $T2$. One of the most persuasive advocates for such properties is Bonjour [Bonjour 85]. We [Kwok, et.al. 98] used these informal properties as the basis of a quantitative approach to measuring the coherence of a theory. As we will be providing a critique of, and an improvement on our earlier approach, we will simply say here that it identified highly coherent theories with those whose formulas occur frequently in supporting observations; thus, we may paraphrase that approach as "coherence = high average use".

The structure of the paper is as follows. In section 2 we outline some traditional arguments against theoretical terms. Section 3 has an example of how one might show that a theoretical term is needed. Support sets are introduced in section 4; there we also critique our earlier definitions in Kwok et.al. (op.cit.). Section 5 defines utility and coherence, and section 6 applies these to some well-known examples. The concluding section 7 summarizes the main points, exposes limitations and suggest further developments of this approach. While most of the formal claims and propostions are quite straightforward, we supply in-line proofs of some of them; however we include an appendix that contains proofs of the less obvious ones.

[1] This means they account for the same set of observations.

2 Science without Theoretical Terms?

Historically, there were two disturbing formal arguments for the possible elimination of theoretical terms from scientific laws. They were predicated on certain formulations of theories, respectively by Ramsey [Ramsey 31] and by Craig [Craig 53]. However, one should be careful not to attribute to them the arguments against theoretical terms as they did not use their formulations in that way. We will briefly review these arguments as our approach to coherence will be used later to examine them in some detail.

Ramsey's method for eliminating theoretical terms is related to Russell's [Russell 56] notion of definite descriptions. Russell desired to provide semantics for terms that do not refer, as in his classic example sentence "The present king of France is bald", where the noun phrase has no contemporary referent. He proposed a formalization of the sentence in logic using existential quantifiers, as in $\exists X \; King(X)$ together with (sub) formulas which express those other properties of X. Hence, if one regards the "The present king of France" as a theoretical term, the overall formula has eliminated any direct use of it. This is the intuition behind *Ramsification*, but Ramsey took it further to also encode the role of the theoretical term. It eliminates a theoretical term t by using an existential quantifier over a variable V together with other (sub) formulas Δ in such a way that any instantiation for V is unique, and moreover Δ identifies the role of this instantiation with that of t.

Craig's method[1], on the other hand, allegedly shows that it is never necessary to use theoretical terms.

Craig's Theorem: Let T be a recursively enumerable theory. Then T has a recursive axiomatization using the language of T.[2]

The alleged (not by Craig!) application of this to the elimination of theoretical terms is as follows. Intepret T as simply all the possibile observation sentences $O_1, O_2, \ldots$, which are intuitively recursively enumerable. Then there is a recursive axiomatization B using only these sentences from which every O_i can be derived. Hence B is a theory for $O_1, O_2, \ldots$ without theoretical terms.

We recall the proof of Craig's theorem because we need it for our later exposition of coherence.

Proof of Craig's Theorem:
From the sequence $O_1, O_2, \ldots$, form the set $B = \{O_1, O_2 \wedge O_2, \ldots, \underbrace{O_i \wedge O_i \wedge \ldots \wedge O_i}_{i\ copies}, \ldots\}$.

Then $B \models O_n$ for each n. Further, if $B \models \beta$, where β is any logical combination of observation sentences, then β is equivalent to a conjunction of some of $O_1, O_2, \ldots$.

Moreover, B is recursive. For, given α, to see if it is in B, do the following. If α is not in the form $\underbrace{A \wedge A \ldots \wedge A}_{k\ copies}$, reject it. If it is, then begin the enumeration $O_1, O_2, \ldots$, and look for O_k. If A is O_k, accept α, else reject it.

There were cogent rebuttals against the (mis?) interpretation of this theorem as an effective abolition of theoretical terms, e.g., that in scientific practice the set of observations is never completed, and that this is intimately related to the notion of precdictive power of B which is not addressed. We will add to the extant objections our notion of coherence. It will be argued below that both Ramsification and Craig's theorem, when

[1] This has some notoriety, and is therefore sometimes nicknamed *Craig's Trick*.

[2] The reverse implication is often posed as an easy exercise in dovetailing arguments in begining logic.

interpreted as attempts to abolish theoretical terms, fail to produce coherent theories. The merit in using coherence is that it is quantitative and specific.

3 Theoretical Terms and Predicate Invention

Predicate invention in machine learning is a practical example of a *vocabulary expansion* to bring in theoretical terms in knowledge representation. Its literature is extensive, but the various approaches are typified by those of Muggleton and Buntine [Muggleton and Buntine 88], Benerji [Banerji 92], and Quinlan [Quinlan 93]. Careful analysis of the philosophical assumptions of such invention was made by Stahl [Stahl 93] in his overview.

Sometimes the *need* for such an invention can be formally demonstrated, for instance when it can be shown that there is no *finite basis* (i.e. no finite axiomatization) for a theory T. In such a case, when T is recursively enumerable, a well-known result of Kleene [Kleene 52] is that if the language of T can be expanded by adding new predicates, a finite basis can always be found. The catch in Kleene's technique is that these new predicates usually do not have any "ontological significance" beyond encoding the requisites for recursive enumeration, thus failing some fundamental requirements of plausibility enunciated by Stahl (op. cit.).

A more mundane example that appeals to the notion of "experimentation" rather than non-finite axiomatizability is a rudimentary blocks world setting in which the *explicit* vocabulary has only one binary predicate $On(\_,\_)$ with arguments the names of the blocks or the table on which they rest. The only constraint is that no block can be in two distinct places. Figure 3.1 represents block configurations consistent with the (partial) theory $T = \{On(A,B), On(A,C), On(B,Table), On(C,Table), \ldots\}$. Intuitively, we know that these two configurations "expand" T in different ways, say $T1$ for the left and $T2$ for the right; but how can we show that this expansion needs an extended vocabulary? Imagine an action which removes block B from the table; the naive physics of the domain would result in block A remaining on block C in the left configuration, but block A would topple on to the table in the right configuration. While there are many ways to represent actions in logic, to keep things simple we consider a STRIPS-like [Fikes and Nilsson 71] representation using the *contraction* operation of the AGM theory of belief revision [Gardenfors 88]. The relevant postulates for AGM contraction are shown as equations 5.1, 5.2 and 5.3 in section 5. Then the result of the action on the left configuration is $T1' = T1 - \{On(B,Table\}$ and that on the right is $T2' = T2 - \{On(B,Table\}$. But experimentally we know that $On(A,C) \in T1'$ while $On(A,Table) \in T2'$, and hence $T1' \cup T2'$ is inconsistent as it violates the constraint. Moreover from the first AGM postulate (op. cit.) for contraction (viz. equation 5.1, $(\Gamma - \alpha) \subseteq \Gamma$ for any theory Γ and any proposition α), this implies that $T1 \cup T2$ was already inconsistent. Then by the Robinson consistency test (see e.g., [Hodges 97]), there is a formula β such that $T1 \models \beta$ and $T2 \models \neg\beta$. This *implicit* β is the expansion to T which separates $T1$ and $T2$. If we assume that the only predicate that captures the informal notion of "where is a block" is $On(\_,\_)$, then β must involve a new predicate[1]

[1] Intuitively we know in this simple domain what it must minimally involve, viz., the "support" relation, or equivalently the "skew" property of block placement.

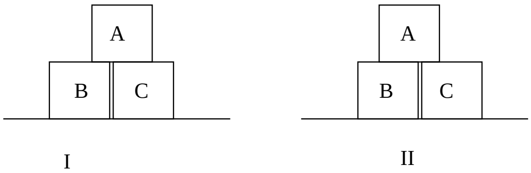

Figure 3.1: The need for a theoretical predicate

4 Support Sets

We now proceed to formalize the notion of coherence. Earlier, we [Kwok, et.al. 98] proposed a way of doing this which attempted to quantify the informal notions mentioned in section 1. Here we retain the general idea of measuring coherence using an averaging technique, but we significantly change what is to be averaged. In [Kwok, et.al. 98] it is the formulas in a theory that account for an observation set that is considered. Here however, we adopt a perspective that is closer to scientific practice.

A simple example will illustrate our view of this practice. The Newtonian equation $F = MA$ can be regarded as a (general) law that, when given as particulars an *input* force f and an *input* mass m will produce the *output* acceleration a. We can choose to regard all of these quantities as observational terms, and the law as a generalization. The use of "input" and "output" does not connote directionality, but merely our choice of which observation terms are used to derive others; in the example above we could have interchanged the roles of force and acceleration. In fact we could go further and posit as an alternative input set f and a, and use the equation to infer m as output.

The formal setting will be a propositional language with later forays into a first-order extension. Thus, we consider a logical language $\mathcal{L}(T)$ of (a scientific) theory T that has two parts: $\mathcal{L}(T)_o$, the *observational* vocabulary, and $\mathcal{L}(T)_t$, the *theoretical* vocabulary. By the *terms* of these vocabularies we will mean their propostional (or predicate) symbols. To re-iterate the distinction made earlier, observational terms are meant to denote entities that can be seen directly, while theoretical terms refer to entities that are only postulated and whose values (truth, etc) may be infered or hypothesized[1] By a theory T we will usually mean a set of formulas and not their logical closure. Therefore, the alternative use of "theory" to signify a logically closed set will correspond to $Cn(T)$ in our notation, where Cn is the logical consequence operator[2]. As suggested by our example above, two subsets I and O of formulas denote respectively the *inputs* and *outputs* of an experimental setting. While we place no restriction on their sub-languages, I will be used as inputs to T and O will be the logical consequences thereof. In predictive uses of T, the set I would comprise mainly observational terms and some hypothesized theoretical terms; the output O will be observational terms to be verified. In abductive uses of T, I will be observational and used with T to produce O which may be partly observational and partly theoretical (i.e. inferring unobservable values). In engineering designs (e.g., band-pass filters) T is circuit theory, I is the performance specification (observables), and O will be the component values (also observables). Other experimental settings can be similarly modelled.

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

[1] Any discomfort with this distinction confirms the misgivings of van Fraassen (op.cit.) and Sellars (op.cit.); clearly the boundary is at least technology dependent.

[2] T is a *base* or *basis* for $Cn(T)$.

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

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

This definition differs from our previous one in [Kwok, et.al. 98] in the relativisation here of the notion of support to the input set I. We believe that this accords with the ordinary use of scientific laws, e.g, by interpreting I as initial or boundary conditions, hypothetical values, measurements, etc., and O with predictions, confirmations, etc., using the laws in T. Thus, different choices of which observations are to be used as starting points will result in different support sets. As limiting cases of the definition, consider the following. If the experimental situation is trivialised by letting I be identical to O (informally, no theory is necessary), then $S(T, I, O)$ is the singleton set $\{\emptyset\}$. If I is empty, then $S(T, \emptyset, O)$ comprises the different economical ways in which T alone can be used to account for O, a kind of oracle. In this way the coherence approach developed here is independent of any commitment to causality or particular use of laws (or rules). Happily, this generalization (and its later ramifications)of support sets preserves most of the results in our earlier paper with only minor modifications, while suggesting novel directions and interpretations.

Observation 1 (Abbreviating $S(T, I, O)$ to $S(T, O)$) *In any context where there is a* fixed *input set I it can be omitted in most discussion, and we can then abbreviate $S(T, I, O)$ to simply $S(T, O)$, and likewise $C(T, I, O)$ to $C(T, O)$ when this does not cause confusion.*

Observation 2 *The set O of (some) of the observational consequences of (base) T therefore need not appear explicitly in T. On the other hand, support sets may sometimes also contain terms from $\mathcal{L}(T)_o$, as these may be explicitly in T and support other observation terms (with the assistance of theoretical terms).*

The support relation, or notions similar to it, occur in many AI areas like diagnoses [Reiter 87], argumentation [Dung 95] and abduction [Denecker and Kakas 02], an unsurprising fact since they each formalize a version of the "account for (fault, conclusion, explanation)" idea in their respective domains. However, definition 1 differs from theirs in making the I-O pair explicit; moreover, we use the support sets $S(T, I, O)$ *combinatorially* by counting the occurrence of certain formulas in them. To keep the counting honest we will need the following assumption.

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

The reason for this is as follows. The single formula $\alpha \wedge \beta \wedge \gamma$ is logically equivalent to the separate formulas α, β and γ. Hence, any finite basis is logically equivalent to a single formula. So if we wish to count the number of formulas in (the basis of) a theory, unless we split conjuncts, no sensible counting can be made.

Despite the simplicity of definition, support sets have interesting properties. Some of them, which easily follow from the definition and compactness, are:

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 $\Gamma_1 \in S(T, I_1, O)$ there is a $\Gamma_2 \in S(T, , I_2, O)$ such that $\Gamma_2 \subseteq \Gamma_1$.

$S(T, I, O)$ is not always as well-behaved as we might hope, as shown by the next observation.

Observation 3 *Suppose $\Gamma_1 \in S(T, I, O_1)$ and $O_1 \subseteq O_2$. Let Δ be a minimal addition to Γ_1 such that $\Gamma_1 \cup \Delta \cup I \models O_2$.*

In general, it is not *the case that $\Gamma_1 \cup \Delta \in S(T, I, O_2)$. The reason is that there may be a proper subset $\Gamma \subset \Gamma_1$ such that $\Gamma \cup \Delta \cup I \models O_2$.*

This is the case when $\Delta \cup \Gamma \cup I \models \alpha$ for some α in $\Gamma_1 \backslash \Gamma$.

5 Utility and Coherence

Within the support sets in $S(T, I, O)$, some formulas (typically, laws or rules) may occur more often than others. These formulas are more frequently used in accounting for O (relative to I), and hence we can ascribe to them a higher "value". A way to measure this value is provided by the notion of *utility* which is defined in the next subsection. A quantification of *coherence* is then based on utility.

5.1 Basic Definitions and Properties

In this subsection we will define *utility* and use it to measure the *coherence* of a theory. Equivalent definitions of coherence are also given.

Definition 2 (Utility of a Formula) *The* Utility *of a formula α in a theory T with respect to an I-relative observation set O is:*

$$U(\alpha, T, I, O) = \frac{\mid \{\Gamma : \alpha \in \Gamma \; and \; \Gamma \in S(T, I, O)\} \mid}{\mid S(T, I, O) \mid} \quad if \quad S(T, I, O) \neq \emptyset$$

Informally, this is the *relative frequency* of occurence of α in the support sets for O. When $S(T, I, O)$ is empty, as it stands above $U(\alpha, T, I, O)$ is undefined. However, to avoid the inconvenience of isolating this case whenever utility is used, we *define* it to be 0 whenever $S(T, I, O)$ is empty. This is not altogether arbitrary; in fact it accords with the relative frequency intuition — if there are no support sets for O, then α is not used in any support set.

As $0 \leq U(\alpha, T, I, O) \leq 1$, the closer it gets to 1 the "more essential" it is. It is 0 if and only if it is irrelevant to the support of O. It is 1 if and only if it is indispensible to the support of O. The latter property is useful enough to record as:

Observation 4 *$U(\alpha, T, I, O) = 1$ if and only if $\alpha \in \Gamma$ for every $\Gamma \in S(T, I, O)$.*

Utility provides a way to realize the informal property of coherence mentioned in section 1, viz., that theories in which the (law-like) formulas are "more useful" are more coherent. We explain this by considering the AGM theory of belief revision

[Gardenfors 88] in which one of the operations is *contraction.* Given a theory T, contracting a formula α from it should result in a minimally changed theory denoted by $T - \alpha$[1]. The essential properties of the contraction operation that we need are captured by the three AGM postulates:

$$(T - \alpha) \subseteq T \tag{5.1}$$

$$\alpha \notin (T - \alpha) \tag{5.2}$$

$$T \subseteq Cn[(T - \alpha) \cup \alpha] \tag{5.3}$$

In the following proposition $\Gamma|_o$ denotes the projection of Γ onto its observational terms.

Proposition 1 *For a finite base T and observation set O in which each observation has a unique support set:*
$Cn(T - \alpha)|_o \subseteq Cn(T - \beta)|_o \Rightarrow U(\alpha, T, I, O) \geq U(\beta, T, I, O)$

This accords with the intuition that if removing a formula α damages a theory more than removing a formula β, then α is "more useful" than β. The condition of unique support sets is a limiting case; less restrictive cases require subtle combinatorial analysis to establish similar results. Without some restrictions, the proposition fails, e.g., when the O-observations that β "covers" overlap extensively while those for α do not.

The coherence definition that follows is similar to Kwok, et.al.'s (op.cit), but again relativised to input observations.

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

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

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. In a sense, the inputs are "free". This in fact accords with scientific and engineering practice where inputs like boundary conditions or design specifications are assumed to be given. The inadequacy of our earlier Kwok, et.al. (op.cit.) definition to allow for the notion of "input" prevented us from applying a potentially fruitful idea to the "good swan fix" example below, and to explaining modularity of theories.

The finiteness assumptions are not essential as the definition (and subsequent results) can be generalized to countable sets using measure theory; however the technicalities of such a generalization may obscure the simplicity of our approach.

Proposition 2

$$0 \leq C(T, I, O) \leq 1 \tag{5.4}$$

$$C(T, I, O) = 1 \leftrightarrow \forall i \forall j \; U(\alpha_i, T, I, O_j) = 1 \tag{5.5}$$

[1] To satisfy postulate 5.2 more than just α may be removed from T if it has formulas that have α as a logical consequence. To ensure minimal removal is the function of postulate 5.3

Corollary 1 *$C(T, I, O) = 1$ iff $S(T, I, O_j) = \{T\}$ for all j.*

That is, a maximally coherent theory is one in which the unique I-relative support set for O is *the entire theory*. Examples of such theories are the programs which are used to define the *Kolmogorov complexity* [Li and Vitanyi 97] of recursively enumerable sequences. One translation into our formalism is as follows. The recursively enumerable sequence is the output observation set O; the program Π is encoded into logic prgramming clauses C_Π; the input set I is empty. Then by the definition of Kolmogorov complexity, every clause in C_Π is needed to produce (or recognize) O. An alternative translation is to let I be the set of natural numbers, in which case O is the sequential enumeration of the sequence. This leads to a natural question: if a theory T has formulas that are redundant with respect to the I-O pair, what happens to its coherence if the redundant formulas are removed? We address this in the next subsection.

There are alternative (and equivalent) ways to view theory coherence. Here is a useful one which shows that coherence can also be regarded as the *average over all observations of the average size of support sets*. The same setting as in 3 is assumed.

Proposition 3

$$C(T, I, O) = \frac{1}{mn} \sum_{j=1}^{m} \frac{1}{\mid S(T, I, O_j) \mid} \sum_{\Gamma \in S(T,I,O_j)} \mid \Gamma \mid$$

This interpretation of coherence can be thought of as a realization of the idea in section 1 that one of its features is that the laws "work better together". In this sense, a theory with small coherence, and therefore small average support set, is more "disconnected". We shall see examples of this interpretation later. A limiting case of propostion 3 is when the observation set is a singleton:

Corollary 2 *Let $T = \{\alpha_1, \alpha_2, ..., \alpha_n\}$ be a theory and let $O = (O_1)$ contain only one observation set. Then,*

$$C(T, I, O) = \frac{1}{n} \frac{1}{\mid S(T, I, O_1) \mid} \sum_{\Gamma \in S(T,I,O_1)} \mid \Gamma \mid$$

A consequence of proposition 3 is that there is a lower bound for the coherence of a theory if it is not 0. The size of non trivial support sets must be at least 1. This implies that the average size of support sets must be at least 1. Normalising this by the size of the theory shows that the lower bound on coherence is the inverse of the cardinality of the theory.

Lemma 2 *Let $T = \{\alpha_1, \alpha_2, ..., \alpha_n\}$ be a theory and let $\vec{O} = (O_1, O_2, ..., O_m)$ be a finite sequence of observation sets. Suppose that $S(T, I, O_j) \neq \{\emptyset\}$ for every j, $1 \leq j \leq m$. Then, $C(T, I, O) \geq \frac{1}{n}$.*

Consider a finite sequence $O = (O_1, O_2, \ldots, O_n)$ of observation sets. Suppose that each observation set O_j in O contains a single formula α_j. Let $T = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, a theory which merely records observations. If it is the case that no observation set in O is entailed by the other observation sets, then $C(T, I, O) = \frac{1}{n}$. The reason for this is that T contains only one support set for each observation set O_j, *viz.* $\{\alpha_j\}$. Thus, *a theory which merely records observations has minimal positive coherence*. We will revisit this later in a related setting when Craig's Trick is re-examined.

5.2 Redundancy

Intuitively, redundant formulas are those that are not needed in a theory because they can be generated from other formulas.

Definition 4 *Let T be a theory. Say that $\alpha \in T$ is redundant if and only if $Cn(T \setminus \{\alpha\})|_O = Cn(T)|_O$. Furthermore, say that T has redundancy if and only if there exists $\alpha \in T$ such that α is redundant in T.*

This notion of redundancy can be expressed in terms of the utility of theory elements and the observational consequences of the theory. An easy consequence is that an element of a theory is redundant if and only if the utility of the theory element is less than 1 for every observational consequence of the theory.

Lemma 3 *Let T be a theory. T has redundancy if and only if for some $\alpha \in T$ and for every $\gamma \in O,\ U(\alpha, T, I, \{\gamma\}) < 1$.*

The relationship between the coherence of subsets of a theory and the coherence of the theory itself can be established by showing how the removal of elements from a theory alters the set of supports for arbitrary observation sets. It should be clear that the support sets for a theory without a given element are simply the support sets of the original theory which do not contain that element. This is recorded as the next lemma, which is also depicted as figure 5.1.

Lemma 4 *Let T be a theory, O an observation set, and let α be an element of T. Then $S(T \setminus \{\alpha\}, I, O) = \{\Gamma \mid \Gamma \in S(T, I, O)$ and $\alpha \notin \Gamma\}$.*

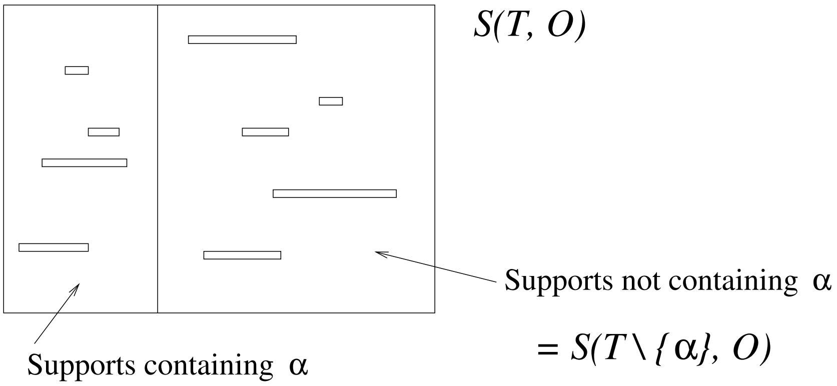

Figure 5.1: Support sets after removing an element

‘A consequence of this is that elements of a theory which have zero utility, relative to certain observation sets, can be removed to increase the coherence of the theory relative to those observation sets. An element of a theory which has a utility of 0 does not appear in any support set. Hence, by lemma 4 the support sets for the contracted theory are exactly the support sets of the original theory. The gain in coherence comes from the decrease in the size of the theory (the average size of the support sets remains unchanged). Examples of theory elements which have zero utility include formulas containing only theoretical terms that are unnecessary for deriving observational consequences. When such formulas are used to generate incompatible but empirically equivalent theories, propostion 4 shows that they can be safely removed to increase coherence. This means that coherence can be used to handle certain simple cases of underdetermined theories.

Proposition 4 *Consider a finite base T and input-output, $I = (I_1, I_2, \ldots, I_m)$ a sequence of (input) observations and $O = (O_1, O_2, \ldots, O_n)$ a sequence of (output) observations for which each (I_j, O_j) pair has a unique support set. If T contains redundancy (i.e., for some $\alpha \in T$, $Cn(T) = Cn(T \setminus \{\alpha\})$) then, for some $\beta \in T$, $U(\beta, T, I_j, O_j) = 0$ for all j and $C(T \setminus \{\beta\}, I, O) = \frac{n}{n-1} C(T, I, O)$.*

A sufficient condition for the existence of a theory element with zero utility is that the theory contains redundancy and that, for the observation sets under consideration, the set of supports of the theory for each observation set is a singleton. Such a restriction is quite strong and it may be expected that theories often contain a number of minimal proofs for consequences.

Lemma 5 *Let T be a theory and $= (O_1, O_2, ..., O_m)$ be a finite sequence of observation sets. If T has redundancy and for every $j, 1 \leq j \leq m, \mid S(T, I, O_j) \mid = 1$, then there exists $\alpha \in T$ such that $U(\alpha, T, I, O) = 0$.*

It would seem intuitive that removing redundant formulae should increase the coherence of a theory. However, the removal of redundant elements from a theory can, in fact, decrease the coherence of a theory. This is because redundant elements can be part of relatively large support sets for an observation set. For example, consider the following theory and observation set:

$T = \{a, \neg c \vee b, c\}$
$O = \{a \vee b\}$

The two support sets of T for O are $\{a\}$ and $\{\neg c \vee b, c\}$. The utility of all three elements in the theory is $\frac{1}{2}$. If c is removed, the utility of $\neg c \vee b$ would fall to 0 while the utility of a would increase to 1 because the only support set of $T \setminus \{c\}$ is $\{a\}$. However, at the extreme values of utility, 0 and 1, removal of theory elements leaves the utility unchanged.

Lemma 6 *Let T be a theory and O an observation set. Suppose $\alpha, \beta \in T$ and $\alpha \neq \beta$. If $U(\alpha, T, I, O) = 0$, then $U(\alpha, T \setminus \{\beta\}, I, O) = 0$. Also, if $U(\alpha, T, I, O) = 1$ and $U(\beta, T, I, O) < 1$, then $U(\alpha, T \setminus \{\beta\}, I, O) = 1$.*

Another result about the utility of theory elements is that the existence of an element with a utility lying strictly between 0 and 1 is sufficient to guarantee the existence of another distinct element of the theory with a utility which is also strictly between 0 and 1. This suggests that theories may contain elements which can be removed to increase the coherence of a theory.

Lemma 7 *Let T be a theory and O an observation set. Suppose $\alpha \in T$ and $0 < U(\alpha, T, I, O) < 1$ then there exists a $\beta \in T, \alpha \neq \beta$ such that $0 < U(\beta, T, I, O) < 1$.*

The coherence of two theories, with one theory being generated from the other by removing elements, can be related to each other. By Lemma 4, the support sets of the truncated theory are simply the support sets of the original theory which do not contain any of the removed elements. Recall that $S(T, I, O)$ is the set of support sets for a theory T and an observation set O, given input I. This notation can be extended to filter out the support sets which contain a certain element of the theory. Thus $S(\alpha_i, T, I, O)$ consists of the support sets in $S(T, I, O)$ which contain α_i.

Definition 5 *Let T be a theory and O an observation set. The α supports of T for O with input I, denoted by $S(\alpha, T, I, O)$, is defined by*

$$S(\alpha, T, I, O) = \{\Gamma \in S(T, I, O) \mid \alpha \in \Gamma\}$$

The relationship between an original theory and a truncated one, generated from the original theory by removing an element with a utility strictly less than 1, is given by the next proposition.

Proposition 5 *Let $T = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a theory, O an observation set, and let $\alpha \in T$. If $U(\alpha, T, I, O) < 1$, then*

$$C(T \setminus \{\alpha\}, I, O) = C(T, I, O) + \\ \frac{(C(T, I, O)((n-1)U(\alpha, T, I, O) + 1)) - \frac{1}{|S(T,I,O)|}\sum_{\Gamma \in S(\alpha,T,I,O)} \mid \Gamma \mid}{(n-1)(1 - U(\alpha, T, I, O))}$$

The constraint that the utility of the element removed is less than 1 is necessary because the removal of an element with a utility of precisely 1 would eliminate all support sets and the truncated theory would no longer entail the observation set. While the above mathematical expression is rather complex, there are basically two interacting factors which give the result. The first is that the original theory has a size of n and the truncated theory has a size of $n-1$. Since the coherence measure is always normalised with respect to the size of the theory, the truncated theory has a multiplicative factor gain of $\frac{n}{n-1}$. The second factor is the average size of support sets for the truncated theory. This is determined by the support sets of the original theory which do not contain the element that is removed. The average size of the support sets of the original theory which do not contain α may be more or less than the overall average size of the support sets of the original theory. The interplay of these two factors give the relationship as described by proposition 5. Notice that the result means that the removal of a non-essential theory element (a theory element with a utility which is strictly less than 1) can either increase or decrease the coherence of a theory. To see how, consider the following theory and observation set:

$$T = \{a, b \vee \neg c, c \vee \neg d, d\}$$
$$O = \{a \vee b\}$$

The set of supports of T (with input I) for O is:

$$S(T, I, O) = \{\{a\}, \{b \vee \neg c, c \vee \neg d, d\}\}.$$

All theory elements are not essential for the observation set O and the removal of different elements of the theory can either result in an increase or a decrease in coherence. Notice that $S(T, I, O)$ contains one large and one small support set. Removing an element of T either eliminates the small or the large support set. When the large

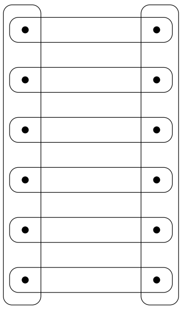

Figure 5.2: A theory for which the removal of any element decreases the coherence.

support set is left, coherence increases. Otherwise, the small support set is left and coherence decreases. The coherence of T for O is $\frac{1}{2}$ since the average size of support sets is 2 and T has 4 elements. Removing a would increase the coherence to 1 because the remaining elements are all contained in the one remaining support set. The removal of any other theory element would decrease the coherence to $\frac{1}{3}$ because the only support set is $\{a\}$ and there are three elements in the contracted theory.

Lemma 7 shows that, if a theory contains an element with a utility in the open interval between 0 and 1, then that theory contains at least two elements which have a utility which is strictly between 0 and 1. While it is always possible to obtain a coherence of 1 for an individual observation set, by adopting one of the support sets as the theory, it is still of interest to consider if and when theory elements can be removed to increase coherence. In general, this is not the case as shown by the theory and the support sets of the theory depicted in Figure 5.2.

An actual theory and observation set which gives such a picture can be easily constructed by letting each point in the picture be a distinct propositional letter. The observation set which would generate such supports contains one formula consisting of a series of disjunctions. Each disjunct would be the conjunction of the elements in each support set. In this figure, each point represents an element of the theory while the points within each box are the elements of a support set. Since each element in a theory is contained within one small support set and one large support set, the removal of any element in the theory will eliminate one small support set as well as one of the two large support sets. In fact, as the "ladder" increases in length, the utility of half the elements in the theory will essentially have their utility halved by the removal of any element in the theory; the other half of the theory will have unchanged utilities. The theory depicted in Figure 5.2 has a coherence of $\frac{1}{4}$ $(= 0.25)$ while the removal of any element reduces the coherence to $\frac{8}{33}$ $(\simeq 0.2424)$.

Despite this example, there are a number of quite general sufficient conditions which imply the presence of a theory element which can be removed to increase coherence. The first result shows that the presence of a "disjoint cover", consisting of non-essential elements of a theory which generate a partition of the support sets, is a sufficient condition. A subset of a theory is a disjoint cover if all support sets contain an element in the subset and for any pair of elements in the subset, α and β, the intersection of the set of support sets which contain α and the set of support sets which contain β is empty. To illustrate this condition, consider Figure 5.3 which depicts the support sets of a theory T for a certain observation set O.

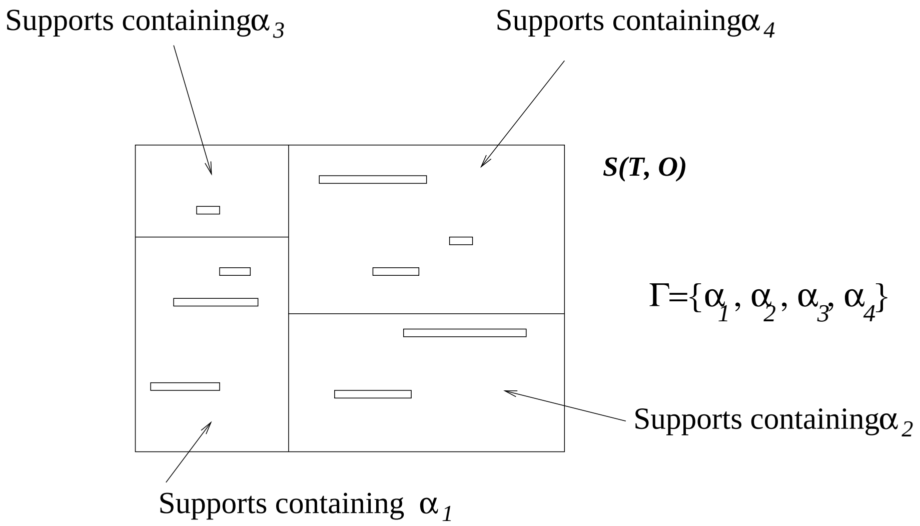
Supports containingα_3
Supports containingα_4
S(T, O)
$\Gamma=\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$
Supports containingα_2
Supports containing α_1

Figure 5.3: Depiction of the support sets of a theory and a disjoint cover Γ of the support sets.

In the figure, the subset Γ of T generates a disjoint partition of $S(T, O)$. Each support set in $S(T, O)$ contains one and only one element of Γ. The condition that a subset of the theory, which is a disjoint cover of the support sets, contains more than one element is sufficient to guarantee that all elements of the cover have a utility which is strictly less than 1. At least one compartment in the partition of the support sets must have support sets which have an average size which is less than or equal to the overall average size of the support sets. This implies that the removal of the theory element in the cover which corresponds to that compartment will leave support sets with an average size that is at least as great as that of the supports for the original theory.

Proposition 6 *Let $T = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a theory and O an observation set. Suppose that for some $\Gamma \subseteq T$,*

1. $|\Gamma| > 1$
2. $S(T, I, O) \neq \emptyset$
3. $S(\alpha, T, I, O) \neq \emptyset$ *for every* $\alpha \in \Gamma$
4. $\bigcup_{\alpha \in \Gamma} S(\alpha, T, I, O) = S(T, I, O)$
5. *for every* $\alpha, \beta \in \Gamma$*, if* $\alpha \neq \beta$ *then* $S(\alpha, T, I, O) \cap S(\beta, T, I, O) = \emptyset$

Then, for some $\alpha \in \Gamma$, $C(T \setminus \{\alpha\}, I, O) \geq \frac{n}{n-1} C(T, I, O)$.

Another sufficient condition which guarantees the existence of a theory element which can be removed to increase the coherence is that if the singleton subset $\{\alpha\}$ is one of the support sets for some theory element α. This implies that all other support sets do not contain α and since all support sets must contain at least one element, removal of α will leave support sets with an average size at least as great as the average size of support sets in the original theory.

Proposition 7 *Let T be a theory and O an observation set. Suppose for some $\alpha \in T, \{\alpha\} \in S(T,I,O)$ and $U(\alpha,T,I,O) < 1$. Then, $C(T\backslash\{\alpha\},I,O) \geq \frac{n}{n-1}C(T,I,O)$.*

A final sufficiency condition which implies increased coherence by removing an element of a theory is if the size of all support sets of the theory for a certain observation set are of the same size. Though an intuitively improbable condition, such a constraint implies that the average size of the support sets remains the same after a non-essential element has been removed.

Proposition 8 *Let T be a theory and O an observation set. Suppose for every $\Gamma, \Gamma^{'} \in S(T,I,O)$ $(\mid \Gamma \mid = \mid \Gamma^{'} \mid)$. For any $\alpha \in T$, if $U(\alpha,T,I,O) < 1$, then $C(T \setminus \{\alpha\},I,O) = \frac{n}{n-1}C(T,I,O)$.*

5.3 Union of Theories

A number of major achievements in science have coincided with the formulation of theories which unify and explain a number of disparate fields of study. These scientific achievements range from the discovery of the connection between magnetism and electricity to the formulation of plate tectonics theory. Coinciding with the emerging study of the chemistry of gases, the oxygen theory of combustion unified the study of gases and combustion. Prior to the formulation by Wegener of plate tectonics theory, a vast number of observations were made which could not be unified or explained. For instance, the layers of rock strata in the east of South America and the west of Africa match very closely. Furthermore, the placement of volcanoes and earthquake zones around the world seemed to follow lines on the earth's surface. Yet, no theory could explain their positions. Since scientific revolutions tend to unify a number of disparate fields, it is of interest to consider how the coherence measure presented here behaves when theories are joined together.

Formulations of scientific theories often contain extralogical axioms which are all necessary before any observational consequences can be derived. This would make the coherence of the theory, as measured here, close to one. For instance, the kinetic theory of gases makes a number of assumptions concerning the behaviour of gas molecules which are all necessary. These assumptions include that molecules collide with each other and with the walls of a container in a perfectly elastic manner. Also, pressure is assumed to be due to the collisions of molecules with the walls of a container. The removal of any of the assumptions would prevent the derivation of any of the observational gas laws.

The coherence measure presented here also plays a role in showing that the resulting theory constructed from the set union of two unrelated theories has a coherence which is certainly less than the maximum of the coherence of the two initial theories. For simplicity in this section we will assume that the input set I is common to both the component theories that will be combined. It is not hard to generalize to the case where each component theory has its own input set, and indeed this is discussed in section 6.4 where the decrease in coherence in the union of some theories is explained intuitively.

As a starting point, it can be shown that the support sets for the union of two theories is a superset of the union of the support sets for each of the two theories.

Lemma 8 *Let T_1 and T_2 be theories, I and O be input and observation sets. Then, $S(T_1 \cup T_2,I,O) \supseteq S(T_1,I,O) \cup S(T_2,I,O)$.*

This shows that the support sets for individual theories are also support sets for the union of the individual theories. Moreover, the union of a number of theories may form new support sets which contain elements from a number of the original theories. However, for two theories with languages having an empty intersection (and thus with observational sub–languages having an empty intersection), the support sets for the union of individual theories coincides exactly with the union of the sets of supports for the individual theories. This happens when two disparate fields of study are unified by simply joining the two areas together without unifying principles. For example, the languages used to describe fossil similarity and earthquake epicentres have no intersection.

Lemma 9 *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be the respective languages for two theories, T_1 and T_2, and suppose that $\mathcal{L}_1 \cap \mathcal{L}_2 = \emptyset$. Then, for input set I and any observation set $O \subseteq \mathcal{L}_1$, $S(T_1 \cup T_2, I, O) = S(T_1, I, O)$.*

Notice in the above result that the roles of T_1 and T_2 are asymmetric but totally interchangeable. Consider two theories that do not share common elements of language and have their coherence measured with respect to two finite sequences of observation sets. The following result shows the connection between the coherence of each of the two theories and the coherence of the union of the two theories.

Proposition 9 *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be the respective languages for theories $T_1 = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and $T_2 = \{\alpha_1^{'}, \alpha_2^{'}, \ldots, \alpha_{n'}^{'}\}$ with common input I. Suppose that $\mathcal{L}_1 \cap \mathcal{L}_2 = \emptyset$. Let $O_1 = (O_1, O_2, \ldots, O_m)$ and $O_2 = (O_{m+1}, O_{m+2}, \ldots, O_{m+m'})$ be two sequences of observation sets such that the observation sets of $\vec{O}_1$ and $\vec{O}_2$ are subsets of the observational language of $\mathcal{L}_1$ and $\mathcal{L}_2$ respectively. Then,*

$$C(T_1 \cup T_2, I, O_2 \circ O_1) = \frac{1}{(n+n^{'})(m+m^{'})}(nmC(T_1, I, O_1) + n^{'}m^{'}C(T_2, I, O_2))$$

The relationship between the coherence of two theories and the coherence of the union of the two theories is similar to that between the velocity of a combined object with the initial objects having different masses and velocities. To gain a better understanding of the result, the following corollary considers the effect of combining two theories of equal size which account for the same number of observation sets and have the same coherence in doing so.

Corollary 3 *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be the respective languages for theories, $T_1 = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and $T_2 = \{\alpha_1^{'}, \alpha_2^{'}, \ldots, \alpha_{n'}^{'}\}$ with common input I. Suppose that $\mathcal{L}_1 \cap \mathcal{L}_2 = \emptyset$. Let $\vec{O}_1 = (O_1, O_2, \ldots, O_m)$ and $\vec{O}_2 = (O_{m+1}, O_{m+2}, \ldots, O_{m+m'})$ be two sequences of observation sets such that the observation sets of $\vec{O}_1$ and $\vec{O}_2$ are subsets of the observational language of $\mathcal{L}_1$ and $\mathcal{L}_2$ respectively. Suppose that $n = n^{'}$, $m = m^{'}$, and $C(T_1, I, \vec{O}_1) = C(T_2, I, \vec{O}_2)$. Then,*

$$C(T_1 \cup T_2, \vec{O}_2 \circ \vec{O}_1) = \frac{1}{2}C(T_1, \vec{O}_1)$$

Thus, when two linguistically unrelated theories of equal size and coherence are combined, the coherence is halved. The main reason for this is that the size of the theory is doubled while the average size of support sets remains the same. This corollary is closely related to the *modularization* property outlined in section 6.4.

The following corollary shows that the coherence of the union of two linguistically unrelated theories is strictly less than the maximal coherence of the two theories.

Corollary 4 *Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be the respective languages for theories, $T_1 = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ and $T_2 = \{\alpha_1^{'}, \alpha_2^{'}, \ldots, \alpha_{n'}^{'}\}$ with common input I. Suppose that $\mathcal{L}_1 \cap \mathcal{L}_2 = \emptyset$. Let $\vec{O_1} = (O_1, O_2, \ldots, O_m)$ and $\vec{O_2} = (O_{m+1}, O_{m+2}, \ldots, O_{m+m'})$ be two sequences of observation sets such that the observation sets of $\vec{O_1}$ and $\vec{O_2}$ are subsets of the observational language of $\mathcal{L}_1$ and $\mathcal{L}_2$ respectively. Then,*

$$C(T_1 \cup T_2, I, \vec{O_2} \circ \vec{O_1}) < \max\{C(T_1, I, \vec{O_1}), C(T_2, I, \vec{O_2})\}$$

In general, when two theories are placed together, without the constraint that the languages of the theories have an empty intersection, the coherence of the union of the theories may increase or decrease. Two theories may each contain a number of little utilised elements. When such theories are joined by set union, the little utilised elements from both theories may combine to form a large support set and the coherence of the united theory may increase. As an example, consider the following theories and observation set:

$I = \emptyset$
$T_1 = \{a, b \rightarrow c, c \rightarrow d, d \rightarrow e, e \rightarrow f\}$
$T_2 = \{b, g, h, i, j\}$
$O = \{a \vee f \vee g\}$

The coherence of T_1 with respect to O is $\frac{1}{5}$ and T_2 with respect to O is $\frac{1}{5}$. But the coherence of the union of T_1 and T_2 with respect to O equals $\frac{1}{10} \cdot \frac{7}{3} = \frac{7}{30}$ which is greater than $\frac{1}{5}$.

6 Some Applications

The intuitive appeal of the above coherence definition will now be tested against a few well-known examples.

6.1 Mendelian Inheritance

In the middle of the 19^{th} century, Gregor Mendel [George 75] correctly hypothesised that two independent characters determine the flower colour of common pea plants (*Pisum sativum*). Each character either codes for purple or white. Since the purple character is dominant, the presence of one purple character ensures that a plant's flowers are purple. Furthermore, parents randomly donate 1 character to each offspring. Thus, when a purple plant is self–fertilised, a plant with 2 purple characters will produce only purple offspring while a plant with only 1 purple character will produce 75% purple offspring and 25% white offspring. It is impossible to tell from appearances whether a pea plant with purple flowers contains 1 or 2 purple characters. However, the colour genes of a plant can be sequenced to produce an observable photograph which falls into one of two equivalence classes. Depending on which class it's DNA sequence photograph falls into, a plant is interpreted as having 1 or 2 purple characters. This can be formalised as the theory T below:

$$\forall x\ [pure(x) \wedge purple(x) \rightarrow selfX\_purple(x, 1)] \tag{6.1}$$

$$\forall x\ [\neg pure(x) \wedge purple(x) \rightarrow selfX\_purple(x, 0.75)] \tag{6.2}$$

$$\forall x\ [DNA\_photo(x,1) \rightarrow pure(x)] \tag{6.3}$$

$$\forall x\ [DNA\_photo(x,2) \rightarrow \neg pure(x)] \tag{6.4}$$

In this example, the observable predicates are $DNA\_photo(\_,\_)$, $purple(\_)$, and $selfX\_purple(\_,\_)$. The $purple$ predicate refers the flower colour of a plant, DNA_photo describes the DNA sequencing result, and $selfX\_purple$ to the precentage of purple flowered offspring from self fertilisation experiments. The one theoretical predicate is $pure$. A plant is pure if it has two identical colour characters and impure otherwise.

This theory can be tested by using as inputs the observable details of particalar plants (flower colour and DNA sequence results) and, as outputs, the results of self fertilisation experiments. For instance, consider the input set $I = \{purple(p1), DNA\_photo(p1,1)\}$ and the observation set $O = \{selfX\_purple(p1,1)\}$. Here the coherence will be 0.5 because 2 out of 4 rules in T are used in the support set. Further, this coherence value will remain at 0.5 for other plant experiments.

6.2 Ramsification and Craig's Theorem

Our approach to coherence trivially excludes the use of Ramsification to abolish theoretical terms for the simple reason that the existential quantifier used in that method must be skolemised to convert the theory to clausal form. The effect of skolemization is to *re-introduce* into the theory theoretical terms that Ramsification was meant to banish. While forcing quantifiers to refer to some entity may be philosophically unpalatable, this is common practice in knowledge representation. This is reinforced by a number of critiques, typical of which is Simon and Groen [Simon and Groen 73].

More problematic is Craig's Theorem. We will use coherence as defined here to argue that the trick invented for the proof of the theorem results in a highly *incoherent* theory. To do this we will assume that the observations are *independent* in the formal sense below; informally this amounts to a sequence of experiments in which the result of any one experiment does *not* subsume or *correlate with* another. One way to imagine this is that care has been taken to design experiments economically to yield maximum information.

Consider the theory
$B = \{O_1, O_2 \wedge O_2, \ldots, \underbrace{O_i \wedge O_i \wedge \ldots \wedge O_i}_{i\ copies}, \ldots\}$
which (using Craig's Trick) is supposed to show that no observation terms are needed.

What is its coherence? To answer this, consider *finite initial segments of B*, viz., $B_n = \{O_1, O_2 \wedge O_2, \ldots, \underbrace{O_n \wedge O_n \wedge \ldots \wedge O_n}_{n\ copies}\}$.
What is its coherence, if the output observation set O is $\{O_1, O_2, \ldots, O_n\}$? As this theory is supposed to be "self-contained" no input is required, so let *I* be $\emptyset$. If the O_i's are "independent", then $O_i \not\models O_j$ for any $i \neq j$. In that case $S(B_n, \emptyset, O_i) = \{O_i\}$ for each i, so $\mid \Gamma \mid = 1$ always.

We have: $S(B_n, \emptyset, O_i) = \{O_i\}$ for each $i \leq n$; $\mid \Gamma \mid = 1$ for each support set. Hence, $\sum_{\Gamma \in S(B_n,\emptyset,O_j)} \mid \Gamma \mid = 1$

Therefore, using proposition 3 in this context:

$$C(B_n, \emptyset, O) = \frac{1}{mn} \sum_{j=1}^{m} \frac{1}{\mid S(B_n,\emptyset,O_j) \mid} \sum_{\Gamma \in S(B_n,\emptyset,O_j)} \mid \Gamma \mid$$

we have:

$$C(B_n, \emptyset, O) = \frac{1}{nn} \sum_{j=1}^{n} 1 \times 1 = \frac{1}{n}$$

Asymptotically, B has 0 coherence!

6.3 The Black Swan Fix

The next well-known example is from the philosophy of science literature that impinges on inductive learning. 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 formula 6.5 in the theory T below:

$$\forall x \; swan(x) \rightarrow white(x). \tag{6.5}$$

Notice that T does not have any theoretical terms as we have specified that both the predicates are observational; but T is a *rule* whose components 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 $sw_1, sw_2, \ldots, sw_n$. The output set O consists of Call this fix T_n. The revised rules that replace rule 6.5 are:
(1 sentence)

$$\forall x \; [swan(x) \wedge \bigwedge_{i \leq n} x \neq sw_i] \rightarrow white(x)] \tag{6.6}$$

and ($2n$ sentences)

$$\bigcup_{i \leq n} \{swan(sw_i), black(sw_i)\} \tag{6.7}$$

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 sw_i, 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[1] is invented to summarize the fact that black swans live in Australia, viz., the new theory T' with 2 sentences:

$$\forall x \; [swan(x) \wedge \neg Australian(x) \rightarrow white(x)] \tag{6.8}$$

$$\forall x \; [swan(x) \wedge Australian(x) \rightarrow black(x)] \tag{6.9}$$

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 6.8 and $\neg Australian(c)$, or by the formulas 6.9 and $Australian(c)$. Therefore, irrespective

[1] This is like the *Abnormality* predicate in non-monotonic reasoning.

of the color the support set for each observation has cardinality 2. Suppose there are k_1 white and k_2 black swans in an observation sequence. It is then easy to see that the utilities of (a) 6.8 is 1 for k_1 observations but 0 for the k_2 observations, (b) 6.9 is the dual of the preceding. Therefore from proposition 3 the coherence of this theory for any $k_1 + k_2$ swans is $\frac{1}{2}$.

6.4 Modularization and Coherence

The coherent fix in the swan example illustrates several features of our definition of coherence that invite further investigation. Here we will merely indicate the interesting connection with aspects of current practice in predicate invention. First, we note that the role of the observable term "Australian"[2] served to increase coherence by capturing regularities. In fact there is more to this than meets the eye. 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 6.8 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 6.9 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. In fact, the Mendelian example above shares this feature with the swan example in being a modular theory in having coherence 1 for each module but only $\frac{1}{2}$ overall.

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 Γ 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 Γ with respect to the observation partitions.

7 Conclusion

An improved quantitative measure of theory coherence has enabled its use in modelling different ways in which theories are empoyed in practice, including prediction, explanation and abduction. This measure was applied to a number of well-known problems and was able to account for the desired intuitive results, e.g. theories that were widely considered informally to be incoherent (respectively, coherent) turned out to have low (respectively, high) measures. Moreover, the measure can be used to identify ways to modularize theories. Future extensions include measures of how formulas work together in groups, and how they relate to entropy methods of predicate invention and selection.

8 Appendix A

This appendix has proofs of some of the lemmas and propositions.

[2] We appreciate that initially the concept of *Australian fauna* was theoretical. However it becomes observable once the specimens are localizable in Australia.

8.1 Proof of proposition 1

By the assumption of unique support sets for each observation O_i in O, there is a 1-1 correspondence between support sets in $S(T, , I, O)$ and the observations. Hence we may a support set by the same suffix as the observation it uniquely supports, abbreviating to S_i that which supports O_i. If S_i is not in $Cn(T - \alpha)$, then by $Cn(T - \alpha)|_o \subseteq Cn(T - \beta)|_o$ it is not in $Cn(T - \beta)$ either. By 5.2, α is not in any S_i that is in $Cn(T - \alpha)$, and by 5.3 those S_j missing from $Cn(T - \alpha)$ are precisely those which contain α. Similar remarks apply to $Cn(T - \beta)$. Moreover, the support sets S_i that are in $Cn(T - \alpha)$ are the same (because of uniqueness) as those in $Cn(T - \beta)$. Hence the number of support sets in T which contain α is at least that which contain β, from which the conclusion follows by definition of utility.

8.2 Proof of proposition 3

For any support set Γ in $S(T, I, O)$, let

$$\delta(\Gamma, i) = \begin{cases} 1 & \text{if } \alpha_i \in \Gamma \\ 0 & \text{otherwise} \end{cases}$$

From this it easily follows that:

$$C(T, I, O) = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{\mid S(T, I, O_j) \mid} \sum_{\Gamma \in S(T,I,O_j)} \delta(\Gamma, i)$$

And in turn, by re-writing this expression for $C(T, I, O)$ we obtain:

$$C(T, I, O) = \frac{1}{mn} \sum_{j=1}^{m} \frac{1}{\mid S(T, I, O_j) \mid} \sum_{\Gamma \in S(T,I,O_j)} \sum_{i=1}^{n} \delta(\Gamma, i)$$

Therefore:

$$C(T, I, O) = \frac{1}{mn} \sum_{j=1}^{m} \frac{1}{\mid S(T, I, O_j) \mid} \sum_{\Gamma \in S(T,I,O_j)} \mid \Gamma \mid$$

8.3 Proof of proposition 4

Let $T = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ be a theory such that $|T| \geq 2$ and $\vec{O} = (O_1, O_2, ..., O_m)$ be a finite set of observation sets. Suppose $C(T, I, \vec{O}) > 0$, and $U(\alpha_n, T, I, \vec{O}) = 0$. Then,

$$\begin{aligned}
C(T \setminus \{\alpha\}, I, \vec{O}) &= \frac{1}{(n-1)m} \sum_{i=1}^{n-1} \sum_{j=1}^{m} U(\alpha_i, T, I, \setminus\{\alpha\}, O_j) \text{ by definition 3} \\
&= \frac{1}{(n-1)m} \sum_{j=1}^{m} \frac{1}{\mid S(T \setminus \{\alpha\}, I, O_j) \mid} \sum_{\Gamma \in S(T \setminus \{\alpha\}, O_j)} \mid \Gamma \mid \text{ by proposition 3} \\
&= \frac{n}{n-1} \left(\frac{1}{nm} \sum_{j=1}^{m} \frac{1}{\mid S(T, O_j) \mid} \sum_{\Gamma \in S(T,I,O_j)} \mid \Gamma \mid \right) \text{ by lemma 4} \\
&= \frac{n}{n-1} C(T, I, \vec{O}) \text{ By proposition 3}
\end{aligned}$$

8.4 Proof of lemma 2

By proposition 3,

$$\begin{aligned} C(T, I, O) &= \frac{1}{nm} \sum_{j=1}^{m} \frac{1}{\mid S(T, I, O_j) \mid} \sum_{\Gamma \in S(T,I,O_j)} \mid \Gamma \mid \\ &\geq \frac{1}{nm} \sum_{j=1}^{m} \frac{1}{\mid S(T, I, O_j) \mid} (|S(T, I, O_j)|) \text{ since for every } \Gamma \in S(T, I, O_j),\, |\Gamma| > 1 \\ &= \frac{1}{nm} \sum_{j=1}^{m} 1 \\ &= \frac{1}{n} \end{aligned}$$

8.5 Proof of lemma 3

$(\Rightarrow)$ Let T be a theory and suppose T has redundancy. Then, for some $\alpha \in T$, $Cn(T \setminus \{\alpha\})|_O = Cn(T)|_O$. Let $\gamma \in Cn(T)$, then $T \setminus \{\alpha\} \models \gamma$. By lemma 1 there exists a $\Gamma \subseteq T \setminus \{\alpha\}$ such that $\Gamma \in S(T, I, \{\gamma\})$. Γ does not contain α and thus $U(\alpha, T, I, \{\gamma\}) < 1$ by the remarks following the definition of utility.

$(\Leftarrow)$ Suppose for some $\alpha \in T$, and every $\gamma \in Cn(T)|_O$ $U(\alpha, T, I, \{\gamma\}) < 1$. Since $U(\alpha, T, I, \{\gamma\}) < 1$, there is some support set, $\Gamma \in S(T, I, \{\gamma\})$ such that $\alpha \notin \Gamma$. Clearly, $\Gamma \subseteq (T \setminus \{\alpha\})$. Thus, $T \setminus \{\alpha\} \models \gamma$. Therefore, $Cn(T \setminus \{\alpha\})|_O = Cn(T)|_O$.

8.6 Proof of lemma 4

Let T be a theory, O an observation set, and let α be an element of T. Suppose $\Gamma \in S(T, I, O)$ and $\alpha \notin \Gamma$. This holds iff $\Gamma \subseteq T$ and $\Gamma \cup I \models O$ and for all $\Gamma^{'} \subset \Gamma$ $(\Gamma^{'} \cup I \not\models O)$ and $\alpha \notin \Gamma$. Which is equivalent to saying that $\Gamma \subseteq T \setminus \{\alpha\}$ and $\Gamma \cup I \models O$ and for all $\Gamma^{'} \subset \Gamma$ $(\Gamma^{'} \cup I \not\models O)$. By definition 1, this is equivalent to saying that $\Gamma \in S(T \setminus \{\alpha\}, I, O)$.

8.7 Proof of corollary 4

Let $T = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ be a theory such that $|T| \geq 2$ and $\vec{O} = (O_1, O_2, ..., O_m)$ be a finite set of observation sets. Suppose $C(T, I, \vec{O}) > 0$, and $U(\alpha_n, T, I, \vec{O}) = 0$. Then,

$$
\begin{aligned}
C(T \setminus \{\alpha\}, I, \vec{O}) &= \frac{1}{(n-1)m} \sum_{i=1}^{n-1} \sum_{j=1}^{m} U(\alpha_i, T \setminus \{\alpha\}, I, O_j) \text{ by definition 3} \\
&= \frac{1}{(n-1)m} \sum_{j=1}^{m} \frac{1}{\mid S(T \setminus \{\alpha\}, I, O_j) \mid} \sum_{\Gamma \in S(T \setminus \{\alpha\}, I, O_j)} \mid \Gamma \mid \text{ by proposition 3} \\
&= \frac{n}{n-1} \left(\frac{1}{nm} \sum_{j=1}^{m} \frac{1}{\mid S(T, O_j) \mid} \sum_{\Gamma \in S(T, I, O_j)} \mid \Gamma \mid \right) \text{ by lemma 4} \\
&= \frac{n}{n-1} C(T, I, \vec{O}) \text{ By proposition 3}
\end{aligned}
$$

8.8 Proof of lemma 5

Suppose α is redundant in T. Let O_j be an observation set in $\vec{O}$. Let $S(T, I, O_j) = \{\Gamma_j\}$ since $S(T, I, O_j)$ is a singleton. By lemma 3, since α is redundant, $U(\alpha, T, I, O_j) < 1$. Thus, $\alpha \notin \Gamma_j$ because otherwise, $U(\alpha, T, I, O_j) = 1$ by the remarks following definition 2. Hence, α is not in any support set and $U(\alpha, T, I, \vec{O}) = 0$.

8.9 Proof of lemma 6

By lemma 4 $S(T \setminus \{\beta\}, I, O) \subseteq S(T, I, O)$ Suppose $U(\alpha, T, I, O) = 0$. Then for every $\Gamma \in S(T, I, O)$, $\alpha \notin \Gamma$. Thus for every $\Gamma \in S(T \setminus \{\beta\}, O)$, $\alpha \notin \Gamma$. Thus $U(\alpha, T, I, \setminus\{\beta\}, O) = 0$. Similarly, if $U(\alpha, T, I, O) = 1$, for every $\Gamma \in S(T, I, O)$, $\alpha \in \Gamma$. Therefore, for every $\Gamma \in S(T \setminus \{\beta\}, I, O)$, $\alpha \in \Gamma$. Thus $U(\alpha, T \setminus \{\beta\}, I, O) = 0$.

8.10 Proof of lemma 7

Suppose $\alpha \in T$. Since $0 < U(\alpha, T, I, O)$, there is a $\Gamma \in S(T, I, O)$ such that $\alpha \in \Gamma$. Similarly, since $U(\alpha, T, I, O) < 1$ there is a $\Gamma^{'} \in S(T, I, O)$ such that $\alpha \notin \Gamma^{'}$. Therefore $\Gamma \not\subset \Gamma^{'}$. Also, since both sets are supports, it must be the case that $\Gamma^{'} \not\subset \Gamma$. This means that for some $\beta \in T, \alpha \neq \beta, \beta \in \Gamma^{'}$ and $\beta \notin \Gamma$. Thus $0 < U(\beta, T, I, O) < 1$.

8.11 Proof of proposition 5

$$\begin{aligned}
C(T \setminus \{\alpha\}, I, O) &= \frac{1}{n-1} \frac{1}{|S(T \setminus \{\alpha\}, I, O)|} \sum_{\Gamma \in S(T \setminus \{\alpha\}, I, O)} \mid \Gamma \mid \text{ by definition 3} \\
&= \frac{1}{n-1} \frac{1}{|S(T,I,O)| - |S(\alpha,T,I,O)|} \sum_{\Gamma \in S(T \setminus \{\alpha\}, I, O)} \mid \Gamma \mid \text{ by corollary 5} \\
&= \frac{1}{n-1} \frac{1}{|S(T,I,O)| - |S(\alpha,T,I,O)|} \sum_{\Gamma \in S(T,I,O) \setminus S(\alpha,T,I,O)} |\Gamma| \text{ by lemma 4} \\
&= \frac{1}{n-1} \frac{1}{|S(T,I,O)| - |S(\alpha,T,I,O)|} \left(\sum_{\Gamma \in S(T,I,O)} |\Gamma| - \sum_{\Gamma \in S(\alpha,T,I,O)} |\Gamma| \right) \\
&= \frac{1}{n-1} \frac{1}{|S(T,I,O)|(1 - U(\alpha,T,I,O))} \left(\sum_{\Gamma \in S(T,I,O)} |\Gamma| - \sum_{\Gamma \in S(\alpha,T,I,O)} |\Gamma| \right) \\
&= \qquad \text{by observation 5} \\
&= \frac{n}{(n-1)(1 - U(\alpha,T,I,O))} \left(\frac{\sum_{\Gamma \in S(T,I,O)} \mid \Gamma \mid}{n \mid S(T,I,O) \mid} \right) - \frac{\frac{1}{|S(T,I,O)|} \sum_{\Gamma \in S(\alpha,T,I,O)} \mid \Gamma \mid}{(n-1)(1 - U(\alpha,T,I,O)} \\
&= \frac{n}{(n-1)(1 - U(\alpha,T,I,O))} (C(T,I,O)) - \frac{\frac{1}{|S(T,I,O)|} \sum_{\Gamma \in S(\alpha,T,I,O)} \mid \Gamma \mid}{(n-1)(1 - U(\alpha,T,I,O)} \\
&= C(T,I,O) + \frac{(C(T,I,O)((n-1)U(\alpha,T,I,O) + 1)) - \frac{1}{|S(T,I,O)|} \sum_{\Gamma \in S(\alpha,T,I,O)} |\Gamma|}{(n-1)(1 - U(\alpha,T,I,O))}
\end{aligned}$$

8.12 Proof of proposition 6

By lemma 11 there exists an $\alpha \in \Gamma$ such that

$$\frac{1}{\mid S(\alpha,T,I,O) \mid} \sum_{\Gamma' \in S(\alpha,T,I,O)} \mid \Gamma' \mid \leq \frac{1}{\mid S(T,I,O) \mid} \sum_{\Gamma' \in S(T,I,O)} \mid \Gamma' \mid .$$

Using this and lemma 12 we see that

$$\frac{1}{|S(T,I,O) \setminus S(\alpha,T,I,O)|} \sum_{\Gamma \in S(T,I,O) \setminus S(\alpha,T,I,O)} |\Gamma| \geq \frac{1}{|S(T,I,O)} \sum_{\Gamma \in S(T,I,O)} |\Gamma|$$

By lemma 10 it also obtains that $U(\alpha,T,I,O) < 1$ which means that $C(T \setminus \{\alpha\}, I, O)$ is defined. Hence,

$$\begin{aligned}
C(T \setminus \{\alpha\}, I, O) &= \frac{1}{(n-1) \mid S(T,I,O) \setminus S(\alpha,T,I,O) \mid} \sum_{\Gamma \in S(T,I,O) \setminus S(\alpha,T,I,O)} \mid \Gamma \mid \\
&\geq \frac{n}{n-1} \left(\frac{1}{n \mid S(T,I,O) \mid} \sum_{\Gamma \in S(T,I,O)} \mid \Gamma \mid \right) \text{ by lemma 12} \\
&= \frac{n}{n-1} C(T,I,O)
\end{aligned}$$

8.13 Proof of proposition 7

Since $\{\alpha\} \in S(T, I, O)$, for every $\Gamma \in S(T, I, O), if \Gamma \neq \{\alpha\}$ then $\alpha \notin \Gamma$. Otherwise $\{\alpha\} \subset \Gamma$ contradicting the minimality of support sets. Further, $\mid \Gamma \mid \geq |\{\alpha\}|$. Thus

$$\begin{aligned} C(T \setminus \{\alpha\}, I, O) &= \frac{1}{(n-1) \mid S(T \setminus \{\alpha\}, I, O) \mid} \sum_{\Gamma \in S(T \setminus \{\alpha\}, I, O)} \mid \Gamma \mid \\ &\geq \frac{1}{(n-1)(\mid S(T \setminus \{\alpha\}, I, O) \mid +1)} \left(\mid \{\alpha\} \mid + \sum_{\Gamma \in S(T \setminus \{\alpha\}, I, O)} \mid \Gamma \mid \right) \\ &= \frac{n}{n-1} \frac{1}{n \mid S(T, I, O) \mid} \sum_{\Gamma \in S(T, I, O)} \mid \Gamma \mid \\ &= \frac{n}{n-1} C(T, I, O) \end{aligned}$$

8.14 Proof of proposition 8

For every $\alpha \in T$, $U(\alpha, T, I, O) < 1$ implies that $S(T \setminus \{\alpha\}, I, O) \neq \emptyset$. Further, for any $S \subseteq S(T, I, O)$

$$\frac{1}{\mid S \mid} \sum_{\Gamma \in S} \mid \Gamma \mid = \frac{1}{\mid S(T, I, O) \mid} \sum_{\Gamma \in S(T, I, O)} \mid \Gamma \mid$$

Thus,

$$\begin{aligned} C(T \setminus \{\alpha\}, I, O) &= \frac{1}{(n-1) \mid S(T \setminus \{\alpha\} \mid} \sum_{\Gamma \in S(T \setminus \{\alpha\}, I, O)} \mid \Gamma \mid \\ &= \frac{n}{n-1} \left(\frac{1}{n \mid S(T, I, O) \mid} \sum_{\Gamma \in S(T, I, O)} \mid \Gamma \mid \right) \\ &= \frac{n}{n-1} C(T, I, O) \end{aligned}$$

8.15 Proof of lemma 8

Let T_1 and T_2 be theories and O an observation set. Suppose that $\Gamma \in S(T_1, I, O)$ then by lemma 1, $\Gamma \in S(T_1 \cup T_2, I, O)$. Thus $S(T_1, I, O) \subseteq S(T_1 \cup T_2, I, O)$. Similarly, $S(T_2, I, O) \subseteq S(T_1 \cup T_2, I, O)$. Hence $S(T_1 \cup T_2, I, O) \supseteq S(T_1, I, O) \cup S(T_2, I, O)$.

8.16 Proof of lemma 9

By lemma 8, $S(T_1, I, O) \subseteq S(T_1 \cup T_2, I, O)$. Thus, what is left is to demonstrate that $S(T_1 \cup T_2, I, O)$ is a subset of $S(T_1, I, O)$. Suppose that $\Gamma \in S(T_1 \cup T_2, I, O)$. Then $\Gamma = \Gamma_1 \cup \Gamma_2$ where $\Gamma_1 \subseteq T_1$ and $\Gamma_2 \subseteq T_2$. Further, this decomposition of Γ is unique since $\mathcal{L}_1 \cap \mathcal{L}_2 = \emptyset$.
In the case that $\Gamma_2 = \emptyset$, $\Gamma \subseteq T_1$ and since $\Gamma \in S(T_1 \cup T_2, I, O)$, for every $\Gamma' \subset \Gamma$, $\Gamma' \not\models O$. Thus, by definition, $\Gamma \in S(T_1, I, O)$.
Otherwise, $\Gamma_2 \neq \emptyset$. In this case, let $\alpha_1 = \bigwedge_{\alpha \in \Gamma_1} \alpha$.
Then, since Γ is a support set of $T_1 \cup T_2$ for O, for every $\beta \in O$, $\Gamma_1 \cup \Gamma_2 \models \beta$.

Let β_0 be an arbitrary element of O. Then, $\Gamma_1 \cup \Gamma_2 \models \beta_0$ and $\Gamma_2 \cup \alpha_1 \models \beta_0$ since α_1 is the conjunction of all the elements in Γ_1. Hence, by the deduction theorem, $\Gamma_2 \models \alpha_1 \rightarrow \beta_0$. This implies that $\Gamma_2 \cup \{\neg(\alpha_1 \rightarrow \beta_0)\}$ is not satisfiable. By the Robinson consistency test [Hodges 97] there exists a $\gamma \in \mathcal{L}_1 \cap \mathcal{L}_2$ such that $\Gamma_2 \models \gamma$ and $\neg(\alpha_1 \rightarrow \beta_0) \models \neg\gamma$. Since $\neg(\alpha_1 \rightarrow \gamma) \in \mathcal{L}_1$, this is a contradiction because $\mathcal{L}_1 \cap \mathcal{L}_2 = \emptyset$.
Thus, $S(T_1 \cup T_2, I, O) \subseteq S(T_1, I, O)$ and hence $S(T_1 \cup T_2, I, O) = S(T_1, I, O)$.

8.17 Proof of proposition 9

$$\begin{aligned}
C(T_1 \cup T_2, I, \vec{O_2} \circ \vec{O_1}) &= \frac{1}{(n+n^{'})(m+m^{'})}\left(\sum_{j=1}^{m+m^{'}} \sum_{\Gamma \in S(T_1 \cup T_2, I, O_j)} \frac{|\Gamma|}{|S(T_1 \cup T_2, I, O)|}\right) \\
&\quad \text{by definition 3} \\
&= \frac{1}{(n+n^{'})(m+m^{'})}\left(\sum_{j=1}^{m} \sum_{\Gamma \in S(T_1 \cup T_2, I, O_j)} \frac{|\Gamma|}{|S(T_1 \cup T_2, I, O)|}\right) + \\
&\quad \frac{1}{(n+n^{'})(m+m^{'})}\left(\sum_{j=m}^{m+m^{'}} \sum_{\Gamma \in S(T_1 \cup T_2, I, O_j)} \frac{|\Gamma|}{|S(T_1 \cup T_2, I, O)|}\right) \\
&= \frac{1}{(n+n^{'})(m+m^{'})}\left(\sum_{j=1}^{m} \sum_{\Gamma \in S(T_1, I, O_j)} \frac{|\Gamma|}{|S(T_1, I, O)|}\right) + \\
&\quad \frac{1}{(n+n^{'})(m+m^{'})}\left(\sum_{j=m}^{m+m^{'}} \sum_{\Gamma \in S(T_2, I, O_j)} \frac{|\Gamma|}{|S(T_2, I, O)|}\right) \\
&\quad \text{by lemma 9} \\
&= \frac{1}{(n+n^{'})(m+m^{'})}\left(nmC(T_1, I, \vec{O_1}) + n^{'}m^{'}C(T_2, I, \vec{O_2})\right)
\end{aligned}$$

8.18 Proof of corollary 3

By proposition 9

$$C(T_1 \cup T_2, I, \vec{O_2} \circ \vec{O_1}) = \frac{1}{(n+n^{'})(m+m^{'})}\left(nmC(T_1, I, \vec{O_1}) + n^{'}m^{'}C(T_2, I, \vec{O_2})\right)$$

Thus, given that $n = n^{'}$, $m = m^{'}$, and $C(T_1, I, \vec{O_1}) = C(T_2, I, \vec{O_2})$ it follows that

$$\begin{aligned}
C(T_1 \cup T_2, I, \vec{O_2} \circ \vec{O_1}) &= \frac{1}{4nm}(2nmC(T_1, I, \vec{O_1})) \\
&= \frac{1}{2}C(T_1, I, \vec{O_1}))
\end{aligned}$$

8.19 Proof of corollary 4

By proposition 9

$$C(T_1 \cup T_2, I, \vec{O_2} \circ \vec{O_1}) = \frac{1}{(n+n^{'})(m+m^{'})}\left(nmC(T_1, I, \vec{O_1}) + n^{'}m^{'}C(T_2, I, \vec{O_2})\right)$$

Without loss of generality, suppose that $C(T_1, I, \vec{O_1}) \geq C(T_2, I, \vec{O_2})$. Then,

$$\begin{aligned} C(T_1 \cup T_2, I, \vec{O_2} \circ \vec{O_1}) &\leq \frac{nm + n^{'}m^{'}}{(n + n^{'})(m + m^{'})} C(T_1, \vec{O_1}) \\ &< C(T_1, \vec{O_1}) \end{aligned}$$

Thus,

$$C(T_1 \cup T_2, I, \vec{O_2} \circ \vec{O_1}) < \max\{C(T_1, I, \vec{O_1}), C(T_2, I, \vec{O_2})\}$$

9 Appendix B

This appendix contains some observations and results that are somewhat peripheral to the main thrust of the paper, but may nevertheless be of interest. A few of them are also cited in the proofs in the preceeding appendix.

Observation 5 *Let T be a theory and O be an observation set. If $\alpha \in T$ then $|S(\alpha, T, I, O)| = U(\alpha, T, I, O)|S(T, I, O)|$.*

9.1 Proof of observation 5

By definition of $S(T, I, O)$:

$$\begin{aligned} U(\alpha, T, I, O) &= \frac{|\{\Gamma \in S(T, I, O) \mid \alpha \in \Gamma\}|}{|S(T, I, O)|} \\ &= \frac{|S(\alpha, T, I, O)|}{|S(T, I, O)|} \text{ by Definition 5} \end{aligned}$$

Thus, $|S(\alpha, T, I, O)| = U(\alpha, T, I, O)|S(T, I, O)|$.

Observation 6 *Let $T = \{\alpha_1, \alpha_2, ..., \alpha_n, \}$ be a theory and O an observation set. Then,*

$$\bigcup_{i=1}^{n} S(\alpha_i, T, I, O) = S(T, I, O).$$

9.2 Proof of observation 6

For each i, $1 \leq i \leq n$, $S(\alpha_i, T, I, O) \subseteq S(T, I, O)$. Thus

$$\bigcup_{i=1}^{n} S(\alpha_i, T, I, O) \subseteq S(T, I, O).$$

Let $\Gamma \in S(T, I, O)$. Let α_i be any element of Γ. Then $\Gamma \in S(\alpha_i, T, I, O)$. Thus

$$\Gamma \in \bigcup_{i=1}^{n} S(\alpha_i, T, I, O).$$

Corollary 5 *Let T be a theory, O an observation set, and let α be an element of T. Then $|S(T \setminus \{\alpha\}, I, O)| = |S(T, I, O)| - |S(\alpha, T, I, O)|$.*

9.3 Proof of corollary 5

By lemma 4, $S(T \setminus \{\alpha\}, I, O) = \{\Gamma \mid \Gamma \in S(T, I, O) \text{ and } \alpha \notin \Gamma\}$. Thus, $|S(T \setminus \{\alpha\}, I, O)| = |\{\Gamma \mid \Gamma \in S(T, I, O) \text{ and } \alpha \notin \Gamma\}|$. Since T is finite, $|\{\Gamma \mid \Gamma \in S(T, I, O) \text{ and } \alpha \notin \Gamma\}| = |S(T, I, O)| - |S(\alpha, T, I, O)|$.

Observation 7 *Let T be a theory and O an observation set. Suppose that for some $\Gamma \subseteq T$,*

$$\bigcup_{\alpha \in \Gamma} S(\alpha, T, I, O) = S(T, I, O)$$

and for every $\alpha, \beta \in \Gamma$, if $\alpha \neq \beta$ then $S(\alpha, T, I, O) \cap S(\beta, T, I, O) = \emptyset$. Then,

$$\sum_{\alpha \in \Gamma} |S(\alpha, T, I, O)| = |S(T, I, O)|$$

9.4 Proof of observation 7

Let $\Gamma = \{\alpha_1, \alpha_2, \ldots, \alpha_l\}$. Using the Inclusion-Exclusion Principle,

$$\begin{aligned}
&|\textstyle\bigcup_{i=1}^{l} S(\alpha_i, T, I, O)| = \\
&\textstyle\sum_{i=1}^{l} |S(\alpha_i, T, I, O)| \\
&-|S(\alpha_1, T, I, O) \cap S(\alpha_2, T, I, O)| - |S(\alpha_1, T, I, O) \cap S(\alpha_3, T, I, O)| - \cdots - |S(\alpha_{l-1}, T, I, O) \cap S(\alpha_l, T, I, O)| \\
&+|S(\alpha_1, T, I, O) \cap S(\alpha_2, T, I, O) \cap S(\alpha_3, T, I, O)| + \cdots + |S(\alpha_{l-2}, T, I, O) \cap S(\alpha_{l-1}, T, I, O) \cap S(\alpha_l, T, I, O)| - \cdots \\
&+(-1)^{l+1}|S(\alpha_1, T, I, O) \cap S(\alpha_2, T, I, O) \cap \cdots \cap S(\alpha_l, T, I, O)|
\end{aligned}$$

Since for every $\alpha, \beta \in \Gamma$, $S(\alpha, T, I, O) \cap S(\beta, T, I, O) = \emptyset$ for every $\alpha \neq \beta$, this expression simplifies to:

$$|\bigcup_{i=1}^{l} S(\alpha_i, T, I, O)| = \sum_{i=1}^{l} |S(\alpha_i, T, I, O)|$$

Further, because

$$\bigcup_{\alpha \in \Gamma} S(\alpha, T, I, O) = S(T, I, O)$$

it follows that $|S(T, I, O)| = |\bigcup_{\alpha \in \Gamma} S(\alpha, T, I, O)|$. Therefore, $\sum_{i=1}^{l} |S(\alpha_i, T, I, O)| = |S(T, I, O)|$.

Lemma 10 *Let T be a theory and O an observation set. Suppose that for some non empty subset, Γ, of T,*

1. *$S(T, I, O) \neq \emptyset$*
2. *$S(\alpha, T, I, O) \neq \emptyset$ for every $\alpha \in \Gamma$*
3. $$\bigcup_{\alpha \in \Gamma} S(\alpha, T, I, O) = S(T, I, O)$$
4. *for every $\alpha, \beta \in \Gamma$, if $\alpha \neq \beta$ then $S(\alpha, T, I, O) \cap S(\beta, T, I, O) = \emptyset$*

Then, $|\Gamma| > 1$ if and only if $U(\alpha, T, I, O) < 1$ for every $\alpha \in \Gamma$.

9.5 Proof of lemma 10

($\Rightarrow$) Suppose that $|\Gamma| > 1$. Let $\alpha \in \Gamma$. Then, since $|\Gamma| > 1$ there exists a distinct $\beta \in \Gamma$. By assumption $S(\beta, T, I, O) \neq \emptyset$ and $S(\alpha, T, I, O) \cap S(\beta, T, I, O) = \emptyset$. Hence $S(\alpha, T, I, O) \subset S(T, I, O)$ and by observation 4, $U(\alpha, T, I, O) < 1$.

($\Leftarrow$) Suppose that $U(\alpha, T, I, O) < 1$ for every $\alpha \in \Gamma$. By assumption, Γ is non-empty. Thus there exists an $\alpha \in \Gamma$. Now, $U(\alpha, T, I, O) < 1$ and by observation 4, there exists a $\Gamma^{'} \in S(T, I, O)$ such that $\alpha \notin \Gamma^{'}$. Since

$$\bigcup_{\alpha \in \Gamma} S(\alpha, T, I, O) = S(T, I, O)$$

there exists a $\beta \in \Gamma$ such that $\alpha \neq \beta$ and $\beta \in \Gamma^{'}$. Thus, $|\Gamma| > 1$.

Lemma 11 *Let T be a theory and O an observation set. Suppose that for some $\Gamma \subseteq T$,*

$$\bigcup_{\alpha \in \Gamma} S(\alpha, T, I, O) = S(T, I, O)$$

and for every $\alpha, \beta \in \Gamma$, if $\alpha \neq \beta$ then $S(\alpha, T, I, O) \cap S(\beta, T, I, O) = \emptyset$. Then, for some $\alpha \in \Gamma$,

$$\frac{1}{\mid S(\alpha, T, I, O) \mid} \sum_{\Gamma^{'} \in S(\alpha, T, I, O)} \mid \Gamma^{'} \mid \leq \frac{1}{\mid S(T, I, O) \mid} \sum_{\Gamma^{'} \in S(T, I, O)} \mid \Gamma^{'} \mid$$

.

9.6 Proof of lemma 11

Suppose the contrary that for every $\alpha \in \Gamma$

$$\begin{array}{llll}
 & \frac{1}{|S(\alpha,T,I,O)|} \sum_{\Gamma^{'} \in S(\alpha,T,I,O)} \mid \Gamma^{'} \mid & > & \frac{1}{|S(T,I,O)|} \sum_{\Gamma^{'} \in S(T,I,O)} \mid \Gamma^{'} \mid \\
\Leftrightarrow & \sum_{\Gamma^{'} \in S(\alpha,T,I,O)} \mid \Gamma^{'} \mid & > & \frac{|S(\alpha,T,I,O)|}{|S(T,I,O)|} \sum_{\Gamma^{'} \in S(T,I,O)} \mid \Gamma^{'} \mid \\
\Rightarrow & \sum_{\alpha \in \Gamma} \sum_{\Gamma^{'} \in S(\alpha,T,I,O)} \mid \Gamma^{'} \mid & > & \sum_{\alpha \in \Gamma} \frac{|S(\alpha,T,I,O)|}{|S(T,I,O)|} \sum_{\Gamma^{'} \in S(T,I,O)} \mid \Gamma^{'} \mid \ \text{summing over } \alpha \in \Gamma \\
\Leftrightarrow & \sum_{\Gamma \in S(T,I,O)} \mid \Gamma \mid & > & \sum_{\Gamma^{'} \in S(T,I,O)} \mid \Gamma^{'} \mid \left(\frac{1}{|S(T,I,O)|} \sum_{\alpha \in \Gamma} \mid S(\alpha, T, I, O) \mid \right) \\
\Leftrightarrow & \sum_{\Gamma \in S(T,I,O)} \mid \Gamma \mid & > & \sum_{\Gamma^{'} \in S(T,I,O)} \mid \Gamma^{'} \mid \left(\frac{1}{|S(T,I,O)|} \mid S(T, I, O) \mid \right) \\
\Leftrightarrow & \sum_{\Gamma \in S(T,I,O)} \mid \Gamma \mid & > & \sum_{\Gamma \in S(T,I,O)} \mid \Gamma \mid
\end{array}$$

Which is a contradiction. Thus for some $\alpha \in \Gamma$

$$\frac{1}{\mid S(\alpha, T, I, O) \mid} \sum_{\Gamma^{'} \in S(\alpha, T, I, O)} \mid \Gamma^{'} \mid \leq \frac{1}{\mid S(T, I, O) \mid} \sum_{\Gamma^{'} \in S(T, I, O)} \mid \Gamma^{'} \mid$$

Lemma 12 *Let T be a theory and O an observation set. Suppose that for some $\alpha \in T$,*

$$\frac{1}{\mid S(\alpha, T, I, O) \mid} \sum_{\Gamma^{'} \in S(\alpha, T, I, O)} \mid \Gamma^{'} \mid \leq \frac{1}{\mid S(T, I, O) \mid} \sum_{\Gamma^{'} \in S(T, I, O)} \mid \Gamma^{'} \mid$$

then

$$\frac{1}{|S(T, I, O) \setminus S(\alpha, T, I, O)|} \sum_{\Gamma \in S(T,I,O) \setminus S(\alpha,T,I,O)} |\Gamma| \geq \frac{1}{|S(T, I, O)} \sum_{\Gamma \in S(T,I,O)} |\Gamma|$$

9.7 Proof of lemma 12

Elementarily,

$$\sum_{\Gamma \in S(T,I,O) \setminus S(\alpha,T,I,O)} |\Gamma| = \sum_{\Gamma \in S(T,I,O)} |\Gamma| - \sum_{\Gamma \in S(\alpha,T,I,O)} |\Gamma|$$

Thus,

$$\begin{array}{ll} & \frac{1}{|S(T,I,O) \setminus S(\alpha,T,I,O)|} \sum_{\Gamma \in S(T,I,O) \setminus S(\alpha,T,I,O)} |\Gamma| \\ = & \frac{1}{|S(T,I,O) \setminus S(\alpha,T,I,O)|} \left(\sum_{\Gamma \in S(T,I,O)} |\Gamma| - \sum_{\Gamma \in S(\alpha,T,I,O)} |\Gamma| \right) \\ = & \frac{1}{|S(T,I,O)|(1-U(\alpha,T,I,O))} \left(\sum_{\Gamma \in S(T,I,O)} |\Gamma| - \sum_{\Gamma \in S(\alpha,T,I,O)} |\Gamma| \right) \text{ by observation 5} \\ \geq & \frac{1}{|S(T,I,O)|(1-U(\alpha,T,I,O))} \left(\sum_{\Gamma \in S(T,I,O)} |\Gamma| - U(\alpha,T,I,O) \sum_{\Gamma \in S(T,I,O)} |\Gamma| \right) \text{ using the assumption} \\ = & \frac{1}{|S(T,I,O)|} \sum_{\Gamma \in S(T,I,O)} |\Gamma| \end{array}$$

Bibliography

[Gardenfors 88] P. Gardenfors. Knowledge In Flux. MIT Press, Cambridge, MA, 1988.

[Banerji 92] R. B. Banerji. Learning theoretical terms. In S. Muggleton, editor, *Inductive Logic Programming*, pages 93–112. Academic Press, 1992.

[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.

[Denecker and Kakas 02] M. Denecker and A. C. Kakas. Abduction in Logic Prorgamming. In *Computational Logic: Essays in Honour of R.A. Kowalski, Pt 1, LNCS 2407, Springer Verlag, 2002.*

[Dung 95] P. M. Dung. An Argumentation-theoretic Foundation for Logic Programming. *Journal of Logic Programming*, 22(2), 151-171, 1995.

[Fikes and Nilsson 71] R.E. Fikes and N.J. Nilsson. STRIPS: a new approach to the application of theorem proving to problem solving. Artificial Intelligence, 2(3-4), pp. 189-208, 1971.

[van Fraassen 80] B. van Fraassen. *The Scientific Image*, pp 14-19, Clarendon Press, Oxford, 1980.

[Hodges 97] W. Hodges. *A Shorter Model Theory*, Cambridge University Press, 1997.

[George 75] W. George. *Gregor Mendel and Heredity*, Priory Press Limited, 1975.

[Kleene 52] S. C. Kleene. Finite axiomatixation of theories in the predicate calculus using additional predicate symbols. In *Memoirs of the American Mathematical Society, 10*. American Mathematical Society, Reading, MA, 1952.

[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.

[Li and Vitanyi 97] M. Li and P. Vitanyi. *An Introduction to Kolmogorov Complexity and Its Applications*, 2nd Ed., Springer Verlag, 1997.

[Muggleton and Buntine 88] S. Muggleton and W. Buntine. Machine invention of first-order predicates by inverting resolution. In *Proceedings of the Fifth International Machine Learning Workshop*, pages 339–352. Morgan Kaufmann, 1988.

[Quinlan 93] J. R. Quinlan. *C4.5: Programs For Machine Learning*. Morgan Kaufmann, 1993.

[Ramsey 31] F. P. Ramsey. *The Foundation of Mathematics and other Logical Essays*. Routledge & Kegan Paul, 1931.

[Reiter 87] R. Reiter. A Theory of Diagnosis from First Principles. *Artificial Intelligence*, 32, 57-95, 1987.

[Russell 56] B. Russell. The Philosophy of Logical Atomism. In *Logic and Knowledge*, pp 117-281, Allen and Unwin, London 1956.

[Sellars 88] W. Sellars. Theoretical Explanation. Reproduced in J.C. Pitt (ed), *Theories of Explanation*, pp 156-166, Oxford University Press, 1988.

[Simon and Groen 73] H. A. Simon and G. J. Groen. Ramsey eliminability and the testability of scientific theories. *British Journal for the Philosophy of Science*, 24:367–380, 1973.

[Stahl 93] I. Stahl. Predicate invention in ILP – an overview. In *Proceedings of the European Conference on Machine Learning*, pages 313–322. Springer–Verlag, 1993.

[Thagard 89] P. Thagard. Explanatory coherence. *Behavioural and Brain Sciences*, 12:435–467, 1989.