Exhaustive Abduction: A Practical Model Validation Tool

Tim Menzies, Windy Gambetta
Artificial Intelligence Laboratory,
School of Computer Science and Engineering,
University of New South Wales, PO Box 1, Kensington, NSW, Australia, 2033
{timm | windy}@cse.unsw.edu.au

ABSTRACT

Models should be able to reproduce the known behaviour of whatever it is they are trying to model. In its most general form, this test is *abduction*; i.e. the generating an internally-consistent scenario that entails some subset of known observations given certain inputs. Exhaustive abduction (EA) is the generation of all such scenarios. EA can be used to verify a model. If all of the known behaviour cannot be found in any of the generated scenarios, then the model must be faulty. Given that abduction is known to be slow, a reasonable pre-experimental intuition is that EA would not be a practical technique for large models. In the study presented here, EAs were executed for a variety of models of different sizes and internal fan-outs. The limits of EA for the current implementation and the studied models implied that EA has some practical utility as a validation tool.

Keywords: validation, abduction, hypothesis testing, qualitative reasoning, neuroendocrinology.

1. INTRODUCTION

Models should be able to reproduce the known behaviour of whatever it is they are trying to model. In its most general form, this test of a model is *abduction*; i.e. the generating an internally-consistent scenario that entails some subset of known observations given certain inputs[1]. Exhaustive abduction (EA) is the generation of all such scenarios.

The *QMOD* project used EA (which they called *hypothesis testing* (HT)) to verify qualitative neuroendocrinological models of glucose regulation. In the original *QMOD* study (which we call HT1) it was found that a glucose model developed from international refereed publications [28] could not reproduce known behaviour. In all, 109 of 343 (32%) of the known data points from six studied papers could not be explained with reference to this model. Of these detected faults, at least one represented an insight into the process of glucose regulation that had been invisible to conventional scientific review process [6, 7].

HT1 was not broad in its scope: it reported one experiment comprising 24 EAs seeking explanations of one to five observations in terms of a single cause over one. In this study, the generality of *QMOD*-style model validation is explored by studying models ranging in number of nodes *N* from 150 to 1250 nodes with average number of children per node *B* of 1 to 10. Section 2 defines this *QMOD* algorithm and its connection to abduction and EA. Section 3 discusses theoretical problems with EA. Section 4 describes the experiments that detected limits to the current EA implementation. These limits seem to be greater than the models we find constructed in the neuroendocrinological domain and some of contemporary KB practice (defined in table 1). The conclusion, therefore, is that EA has some practical utility as a validation tool.

Application	*N*	*B*
mmu	65	7
tape	80	4
neuron	155	4
displan	55	2
DMS-1	510	6

Table 1: *Model size* N *and average fan-out* B *in the and-or graph of real-world expert systems*[2]. *From* [23]. *A practical validation algorithm must work at least for the range 50*<=N<=*510 and 2*<=B<=*7.*

2. *QMOD* = EA = VALIDATION

This section describes the *QMOD*-style validation and its connection to abduction and EA. We begin by adopting the following apparently simple validation algorithm:

---

1 Consider a system with two facts *a* , *b* and a rule *if a then b*. *Deduction* is the inference from *a* to *b*. *Induction* is the process of learning *if a then b* given examples of *a* and *b* occurring together. *Abduction* is inferring *a*, given *b*. Abduction is a plausible inference only since other rules may have concluded *b* using another premise. Hence abduction requires some inference assessment operator. See [2] for a short tutorial introduction. See [20] for an extensive overview. For a formal analysis of abduction, see [1, 11, 27]. For a list of applications, see the conclusion.

2 This sample size should be larger. However, there is very little published information on the size of real-world expert systems. We use table 1 since it is consistent with the author's knowledge engineering experience [18, 19, 24] and the neuroendocrinological models we are aware of.

ALGORITHM 1: *Generate all possible behaviours from a model, then check that the known behaviour can be found amongst the possible behaviours. If not, then the model is faulty.*

Note that this algorithm is silent on the best internal form of the model. It assumes that issues such as (e.g.) the presence of loops, tautologies, redundancies, inconsistencies are secondary to the basic requirement that a model must be able to reproduce the known behaviour of the thing that it is modelling. For algorithms that critique these internal model features, see [23, 29, 30]. EA is closest in internal data structures to the CTMS-validation procedure of [30]. However, CTMS validation assumes that the model testing process can dictate to the environment what test data is to be supplied. Here, we explore domains where data collection is prohibitively expensive (e.g. neuroendocrinology, ecology, and economics) and we must make do with whatever data is currently available. In such data-starved domains, models and inference engine must include a qualitative (i.e. non-numeric) component.

Algorithm 1 is naive for qualitative domains. Consider the links between *a b*, and *c* in the qualitative model of figure 1.

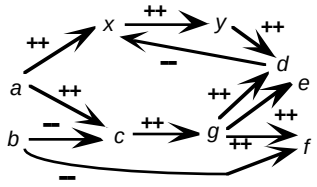

Figure 1: *Connection between entities with legal states* up , down, *or* steady. *Links such as ++ and -- are defined by tables of valid state transitions. For example: (1)* X -- Y *iff* Y *being up or down could be explained by* X *being down or up respectively; (2)* X ++ Y *iff* Y *being up or down could be explained by* X *being up or down respectively.*

In the case of causes $C = \{a\uparrow, b\uparrow\}$[3] we have two competing qualitative influences on *c*: (i) $a\uparrow$ can cause $c\uparrow$ while (ii) $b\uparrow$ can cause $c\downarrow$. Lacking quantitative information about the relative size of these competing forces, one world must be created for each possible value of *c* ; i.e. $\{c\downarrow, c\Theta, c\uparrow\}$ [10]. This branching of behaviours may become computationally intractable and is an unsolved problem in qualitative physics [8], particularly in model with loops (e.g. $\overline{xyd}$ in figure 1 or any model with a control feedback loop). Explanations with loops (e.g *a* went up, then later it went down, then it went up again) are only required for time-variant behaviour. The *C* and *FX* sets from HT1 did not mention time and so HT1 could ban loops from its explanations (heuristic #1). As another technique for reducing the search space across indeterminate models, HT1 also restricted its processing to explanations of known effects resulting from known causes (heuristic #2). We adopt these two heuristics since we will later demonstrate that EA is fundamentally a slow process.

ALGORITHM 2: *Generate all possible non-cyclic behaviours from a (possibly cyclic) model* M *that (i) result from known causes* C*; and (ii) includes some subset of the effects we want to explain* FX. *Check that known behaviour can be found amongst the possible behaviours. If not, then the model is faulty.*

Algorithm 2 contains a bug. Returning to figure 1, consider the case where the effects $FX = \{d\uparrow, e\uparrow, f\downarrow\}$ and causes $C = \{a\uparrow, b\uparrow\}$. Non-cyclic pathways linking *FX* to *C* are:

$$P_1 = \{a\uparrow, x\uparrow, y\uparrow, d\uparrow\} \qquad P_2 = \{a\uparrow, c\uparrow, g\uparrow, d\uparrow\}$$
$$P_3 = \{a\uparrow, c\uparrow, g\uparrow, e\uparrow\} \qquad P_4 = \{b\uparrow, c\downarrow, g\downarrow, f\downarrow\}$$
$$P_5 = \{b\uparrow, f\downarrow\}$$

Note that $P_1 \ldots P_5$ have to make *assumptions* about node values that do not exist in *FX* or *C* (e.g. *b,g,x,* and *y*). While we can find explanations of all of *FX*, some of the explanations use inconsistent assumptions. For example, in a domain of measured continuous variables, *c* and *g* can't be both *up* and *down* in the same scenario. P_2, P_3, and P_4 explain all of *E* but only in the impossible situation that *c* and *g* go up and down simultaneously. Assumptions that contradict assumptions in other paths are called *controversial*.

Algorithm 3 fixes this bug by insisting that the generated explanations are internally consistent.

ALGORITHM 3: *Generate all possible non-cyclic behaviours from a (possibly cyclic) model* M *that (i) result from known causes* C*; and (ii) includes some subset of the effects we want to explain* FX. *Divide these explanations into "worlds": sets of explanatory paths which do not violate domain invariants,* I. *Now see if the known behaviour can be found amongst the possible behaviour subsets. If not, then the model is faulty.*

---

3 Terminology: *C* = a set of causes. *FX* = a set of effects. *M* is a model whose nodes may take one of several mutually exclusive values. M' is a model generated from *M* which has one node for each combination of *M* node/values. The number of nodes and average fanout of M' is *N* and *B* respectively. All members of *C* and *FX* are M' nodes. *I* = model invariants of arity 2 that accept as input M' nodes. *P* = paths from members of *FX* to members of *C* across M' . *W* = worlds = subsets of *P* that do not violate *I*. Lower cases italic letters in the text denote node/value assignments; e.g. $x\uparrow$ = "the value of x has gone up"; $x\downarrow$ = "the value of x has gone down"; $x\Theta$ = "the value of x is steady".

When paths are grouped into worlds, these groups are defined in terms of the *base assumptions*; i.e. the highest controversial assumption in each path. For $P_1 \ldots P_5$ above, we could explore different worlds for each combination of $\{c\uparrow, c\downarrow\} \& \{g\uparrow, g\downarrow\}$. However, since *g* is fully dependant on *c*, *g* will always have the same state as *c* and we can ignore *g* exploring worlds[4]. We therefore have 3 worlds containing those paths that are compatible with $c\uparrow$, $c\downarrow$, and with no value for *c*. That is: $W_0 = \{P_1, P_5\}$[5], $W_1 = W_0 + \{P_2, P_3\}$, $W_2 = W_0 + \{P_4\}$. The cover of a world is the number of effects in it: cover(W_0)= $|d\uparrow, f\downarrow| = 2$; cover($W_1$)= $|d\uparrow, e\uparrow, f\downarrow| = 3$; and cover($W_2$)= $|d\uparrow, f\downarrow| = 2$. We choose to believe the W_1 base assumptions since this permits explanations of the most number of effects. In effect, algorithm 3 has selected a subset of figure 1 (shown in figure 2).

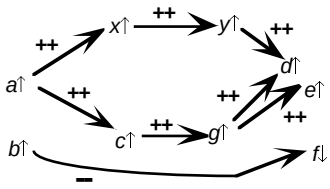

Figure 2: *The subset of figure 1 selected by algorithm 3 that explains the most effects FX ={d↑,e↑,f↓} given causes C ={a↑,b↑}. Note while figure 1 condones inferences between B & C, G & F, and D & X, algorithm 3 has elected not to use these connections.*

Algorithm 3 cannot be viewed in terms of classical deduction which demands that if the rule *x if y* exists, then in every world where *x* is true, *y* is also true [22]. Algorithm 3 is *abductive*: from a space of possible inferences supplied by the user, a subset has been selected according to some criteria in order to fulfil a certain task (in the case of *QMOD*: maximal cover of a set of effects). Indeed, if re-write algorithm 3 in terms of its equivalent logic, we arrive at the formal definitions of abduction as proposed by [5, 21, 27] (see 4.2 of algorithm 4).

ALGORITHM 4 (exhaustive abduction): *4.1) Partially evaluate a FOPC (possibly cyclic) theory* M *w.r.t.. its inference engine to generate* M', *a finite propositional (possibly cyclic) with no negation. 4.2) (abduction) Compute all the non-cyclic models* M" *that satisfy (i)* M" ⊆ M'*; (ii)* M" & C ⊢ (FX$_1$ ⊆ FX)*; (ii)*¬(M" & C ⊢ *false) (i.e. does not violate* I*); and (iii) is maximal (i.e. is not a subset of another* M" *that satisfies (i) and (ii)). 4.3) Pass all the generated* M" *models to an assessment operator* BEST. *If the* BEST *explanations do not completely cover* FX*, then the model is faulty.*

Algorithm 4 partially evaluates *M* to generate *M'* since our current implementation (HT4) uses bit-strings to optimise its internal processing and it assigns one bit to each possible state of each *M* node (which would be distinct *M'* nodes). For example, figure 3 shows the *M'* generated from *M*. In effect, *M'* is the search space tacit in *M*. *M'* must be finite since (i) HT4 will search for all pathways from effects back to causes and (ii) if *M'* is infinite, then this search will never terminate.

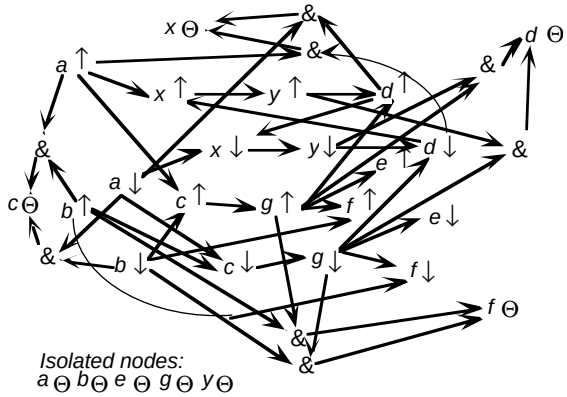

Figure 3: *M' generated from the M of figure 1 assuming that (i) M node can have one of three values: ↑(up), ↓(down,) or Θ(steady); (ii) the conjunction of an up and down can explain a steady; (iii) no change can be explained in terms of a steady (i.e. steady nodes have no children). And-nodes are denoted "&"; for example,* $a\uparrow \& b\uparrow \rightarrow c\Theta$. *All other nodes are or-nodes. Steady nodes that cannot be explained are shown bottom-left.*

Table 2 shows how *N* and *B* can vary between *M* and *M'* models in the neuroendocrinological domain. Note that *N* increases by a very large percent (due to the addition of the and-nodes required for explaining steadies) while the *B* increase is much smaller.

---

4 This approach was inspired by the minimal environment labels of the ATMS [4]. We differ from the ATMS in that we only compute labels for propositions on paths between causes and effects.

5 W_0 denotes the empty set base assumptions world; i.e the world with no controversial assumptions.

	N			B		
Model	*M*	*M'*	*ΔN*	*M*	*M'*	*ΔB*
Figure 1	9	35	389%	1.2	1.26	105%
HT1	80	554	692%	1.7	2.25	130%

Table 2: *Changes in size (*N) *and average fanout (*B) *between M and M' models from the neuroendocrinological domains. Figure 3 shows the M' generated from the* M *of Figure 1.*

How large are the *N* and *B* changes resulting from the *M* to *M'* translation of a propositional rule-base? One node must be created for every literal and its negation. However, given that non-monotonic reasoning is not widely used in commercial practice (e.g. the domains surveyed in Table 1), then only the negated literals inside nested disjunctions and negations in rule premises would be used. For example, the *M* to *M'* translation of *if (a and not (b and c)) or d then e* is shown in figure 4.

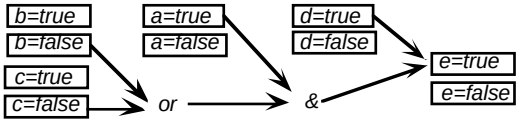

Figure 4: *M' for the M propositional model for* if (a and not (b and c)) or d then e. *Note the isolated negated literals for* a, d *&* e.

We can now define *QMOD*-style validation.

ALGORITHM 5 (QMOD): *Call algorithm 4. Use the following* BEST *operator. Set* MAX_COVER *to the largest size of* FX_1 *in any* M''. *Return all* M'' *models with* MAX_COVER. *The* EXPLAINED *effects are the union of the* FX_1 *of these M'' with maximum cover. If* |EXPLAINED| <> |FX|, *then the model is faulted.*

That is, a model fails *QMOD*-style validation if after making every assumption we could to explain the most number of effects, there remains some inexplicable effects. While *QMOD-BEST* seems weak when the cover of the worlds are nearly the same, it is more reasonable when the coverages are more divergent. For example, in the case where 100 effects can be explained in W_x and only one effect can be explained in W_y, then *QMOD* reports a failure to explain the single effect in W_y. We justify the use of this definition of *BEST* as follows. In terms of theory repair, starting with a partially useful incorrect theory is better than starting with nothing (theory repair in this domain is discussed in [15, 16]).

The implementation of an efficient algorithm 5 is non-trivial. HT1 took 2 days to execute 24 EAs. HT2 & HT3 used a basic chronological backtracking approach (i.e. no memoing) that only terminated for very small models. Basic chronological backtracking has the disadvantage that any feature of the space learnt by the search algorithm is forgotten when backtracking on failure [4, 14]. The current implementation, HT4 learns and caches as much as it can about the search space as it executes. The data required to switch between worlds is also cached so context switching does not require extensive further computation. Also, bit-strings are used to optimise set processing. For more details, see [17].

3. IS EA VALIDATION PRACTICAL?

One pre-experimental pessimistic prediction about EA would be that any behaviour can be generated from a search through indeterminate models. If so, then the validation power of EA would be zero since it would "verify" every model given to it. This is the *Pendrith limit* to EA validation (named after the doctoral student who first succinctly articulated it).

Another pessimistic prediction is that EA validation would be too slow for real-world sized models. EA uses abduction, and abduction is known to be NP-hard [1, 27]; i.e. very likely to be computational intractable in the worst-case. An unfortunate feature of abduction is that this worst-case behaviour is often the usual case: most known abductive inference engines exhibit exponential runtimes for real-world inputs, even for sophisticated algorithms. Hence, many of the articles in [20] are concerned with heuristic optimisations of abduction. Eshghi report a class of polynomial-time abductive inference problems, but this class of problems require a non-cyclic background theory [5]. Bylander reports techniques for tractable abduction [1], but many of these techniques (e.g. rule-out knowledge to cull much of the search space) are not applicable to arbitrary models developed in poorly-measured domains (e.g. neuroendocrinology).

In the case of HT4, the implementation tricks described above (end of §2) do not address the fundamental complexity of the EA task. That is, given the under-specified nature of the models and the exhaustive nature of the inference, the search cannot be culled at the local propagation level. The utility of each local inference has to be assessed by a meta-interpreter using the global criteria: "will it eventually lead to maximal coverage?". This global criteria cannot be applied till after all possible paths are collected (i.e. it cannot be used to cull the search space).

4. EXPERIMENTS WITH HT4

The previous sections motivated the EA algorithm and noted that there were 2 theoretical limits to EA validation: (1) the Pendrith limit to the critiquing power of the process; and (2) a runtime limit due to the possibly

intractable nature of the search. We therefore collected data on how HT4 managed these limits. Two studies were performed comprising 299 models and 4504 EA runs:

1) *The changing model size study:* The HT1 model had 554 nodes in its M' (N=554), an average fanout B of 2.25, and processed 24 EA runs with $|C|$ =1 and 1<=$|FX|$ <= 7. These FX and C size limits were an artefact of certain implementation decisions made the *QMOD* designers; the actual FX and C ranges were 1<= $|FX|$ <= 10 and 1<=$|C|$ <=4. In the changing model size study, nodes were created/ removed to produce 94 models with *150*<=N<=*1250.* Links were added/ deleted to keep the fanout constant at B = 2.25. 1991 EAs for these models were executed for 1<= $|FX|$ <= 10 and 1<=$|C|$ <=4. EA runs that took longer than an arbitrary "give-up" time of 5 minutes were aborted, and that run assigned a time of 5 minutes.
2) *The changing fanout study:* Again, starting with the HT1 model, links were added to produce 205 models with 2 <=B <= 10 and N constant at N=554. 2513 EAs for these models were executed for 1<= $|FX|$ <= 10 and 1<=$|C|$ <=4.

For each study, two graphs were generated: average runtimes and percentage explainable effects. HT4 was to be accepted as a practical validation tool if in the range of 2<=B<=7 and 50<=N<=510 (i) the Pendrith limit was not prohibitive; i.e. the percent explainable was usually less than 100%; and (ii) the runtimes were acceptable (i.e. less than some "too-slow" time which we will set to five minutes).

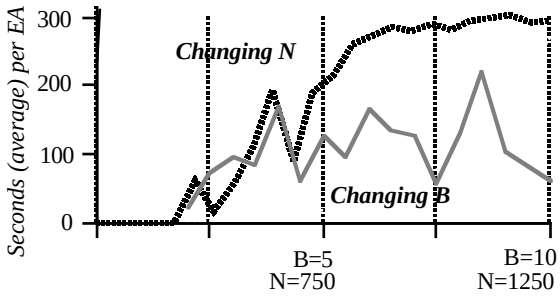

Figure 5: *Runtimes from the changing model size* N *study and changing fanout* B *study. Note that the plateau after* N=*800 is an artefact of the "give-up" time limit of 5 minutes (300 seconds).*

The runtimes shown in figure 5 satisfy our runtime requirements. The variance in the data on the changing N curve prevent a definitive statement regarding the experimentally observed complexity (i.e. exponential or otherwise). Experiments continue to confirm/ refute the exponential nature of HT4's EA inference.

The changing B result of figure 5 is somewhat surprising. The pre-experimental intuition was that runtimes would be exponential on fanout since increasing fanout in a graph containing two nodes X and Y increases exponentially the number of paths between X and Y. The observed B increase was therefore surprisingly small. However, several factors could counter any increase. (i) Frequent incompatibilities of nodes on possible paths would cull the total number of paths generated (i.e. violations of I cull the search space). (ii) Adding links around an and-node increases the pre-conditions to propagation of the search over that and-node. That is, sometimes adding links adds extra constraints which restricts the number of possible paths.

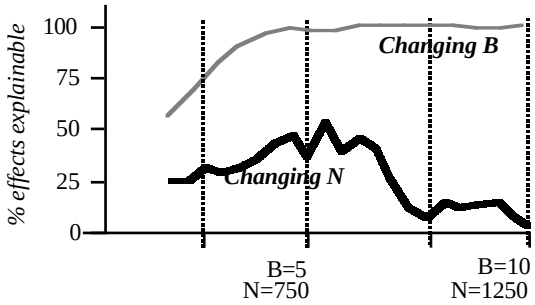

Figure 6: *% effects explainable from the changing model size* N *study and changing fanout* B *study. Figure 5 shows that after* N=*800, most of the runs did not terminate before the "give-up" time of 300 seconds. Hence, the changing* N *curve in this figure drops off suddenly after* N=*800 .*

Figure 6 demonstrates two interesting features of EA search. Firstly, after an average fanout of 4, the Pendrith limit is reached and the model can explain all supplied behaviours. Secondly, for models below the Pendrith limit, between 45% to 75% of the effects from 2513 runs were inexplicable; i.e. EA provides a significant level of critique.

This B=4 Pendrith limit is less than the Table 1 figures of fanout in real-world propositional expert systems. Hence, the Pendrith limit appears to restricts the utility of EA in that domain. However, Table 3 demonstrates that for domains which use a non-simple M to M' translation (e.g. neuroendocrinology) that certain node types may have a fanout much larger than the average fanout. The strength of this observation is that, for the neuroendocrinology domain, M models are built in terms of *changes*; i.e. the domain expert's natural idiom contains phrases like "if X goes up then Y goes down".

To assess the impact of the Pendrith limit on modelling, we should check the average fanout for expert-created M nodes in the M' model. Table 4 shows us that at an average B=4 for the 205 changing fanout models, the

fanout of the expert-created nodes (the changes) was *B*=9; i.e. higher than our Table 1 range. We therefore conclude that the Pendrith limit is not a practical limit to EA in neuroendocrinology.

Node Type	*changes*	*steadies*	*ands*	*all*
E.G.	$a\uparrow, a\downarrow$	$a\Theta$	&	$a\uparrow, a\downarrow, a\Theta, \&$
N = Number of nodes	18	9	8	35
Σ = *Total number of kids*	36	0	8	44
Σ / N = *Average B*	2	0	1	1.26

Table 3: *Average fanout (*B*) for different node types in figure 3. Note that the average total fanout may be different to the average fan out of different node types.*

Node Type	*changes*	*steadies*	*ands*	*all*
Average B	9	0	2.5	4

Table 4: *Average fanout (*B*) for different node types in the changing fanout study at the Pendrith Limit. Note that at an average fanout of 4, the fanout of the expert-created nodes (i.e. the changes) is 9 and therefore greater than our Table 1 boundaries.*

5. CONCLUSION

We have explored the link between knowledge-base validation and abductive inference. Validation-as-abduction has a natural application to any model built in an abductive domain; e.g. model-based diagnosis [3]; natural language processing (see multiple examples in [20]); explanation generation [12]; visual pattern recognition and design [21, 22]; frame-based reasoning [25, 26] and case-based reasoning [13]. Even in domains that are apparently non-abductive (e.g. qualitative reasoning in neuroendocrinology or deduction in propositional rule bases), exhaustive abduction neatly characterises the process of validating that a model can somehow explain known behaviour.

Two disadvantages with characterising validation-as-abduction are: (i) slow runtimes and (ii) the multiple-worlds nature of the inference permitting a possible explanation of any behaviour. In the studies described here, we have seen that the runtimes are not unacceptably slow for the models we see in contemporary practice. However, after an average fanout of 4, validation-as-abduction losses its critiquing power. In practical terms, this implies a limit of the validation-as-abduction of some propositional systems. For domains with a more complex semantics (e.g. qualitative physics), this Pendrith limit appears not to be so critical (see the discussion around table 4).

These conclusions are based on a less-than-optimum sample size. Our definition of "real-world expert systems" comes from a single source (see Table 1). All of the 299 models used in our studies were generated from the internal parameters of a single neuroendocrinological model. We are sensitive to the criticism that this sample size is too small to make a general conclusion. However, we believe that we mutated that model over a sufficiently wide range to claim that our experimental results have some generality.

EA is not the first validation algorithm defined using a multi-worlds logic (see [9, 30] for others). However, to our knowledge, this is the first time that the limits to such an algorithm have been experimentally tested using a large test suite. We note that a lack of readily-available models need not be a restriction to such testing. Any number of models and data sets can be artificially generated using known models/ data sets as a reference point, then changing the parameters as required.

6. REFERENCES

1. Bylander, T., D. Allemang, M.C. Tanner, and J.R. Josephson, *The complexity of abduction* **Artificial Intelligence**, 1991. **49**: p. 25-60.
2. Charniak, E. and D. McDermott, **Introduction to Artificial Intelligence**. 1987, Addison-Wesley. 701.
3. Console, L. and P. Torasso, *A spectrum of definitions of model-based diagnosis* **Computational Intelligence**, 1991. **7**(3): p. 133-141.
4. DeKleer, J., *An Assumption-Based TMS* **Artificial Intelligence**, 1986. **28**: p. 163-196.
5. Eshghi, K. *A Tractable Class of Abductive Problems* in **IJCAI '93**. 1993. Chambery, France:
6. Feldman, B.T., P.J. Compton, and G.A. Smythe. *Towards Hypothesis Testing: JUSTIN, Prototype System Using Justification in Context* in **Proceedings of the Joint Australian Conference on Artificial Intelligence, AI '89**. 1989.
7. Feldman, B.Z., P.J. Compton, and G.A. Smythe. *Hypothesis Testing: an Appropriate Task for Knowledge-Based Systems* in **4th AAAI-Sponsored Knowledge Acquisition for Knowledge-based Systems Workshop**. 1989. Banff, Canada, October 1989.:
8. Fouche, P. and B. Kuipers, *An Assessment of Current Qualitative Simulation Techniques*, , B. Faltings and P. Struss, Editor. 1992, The MIT Press: Cambridge, Mass. p. 263-278.
9. Ginsberg, A. *A new approach to checking knowledge bases for inconsistency and redundancy* in **Proc. 3rd Annual Expert Systems in Government Conference**. 1987. IEEE Computer Society.
10. Iwasaki, Y., *Qualitative Physics*, in **The Handbook of Artificial Intelligence**, A. Barr, P.R. Cohen, and E.A. Feigenbaum, Editor. 1989, Addison Wesley: p. 323-413.
11. Kabas, A.C. and P. Mancrella. *Generalized Stable Models: A Semantics for Abduction* in **ECAI-90**. 1990. Stockholm, Sweden:

12. Leake, D.B., *Goal-Based Explanation Evaluation* **Cognitive Science**, 1991. **15**: p. 509-545.

13. Leake, D.B. *Focusing Construction and Selection of Abductive Hypotheses* in **IJCAI '93**. 1993.

14. Mackworth, A.K., *Consistency in Networks of Relations* **Artificial Intelligence**, 1977. **8**: p. 99-118.

15. Mahidadia, A.J., P. Compton, T.J. Menzies, C. Sammut, and G.A. Smythe. *Inventing Causal Qualitative Models: A Tool for Experimental Research*, in **Proceedings of AI '92**. 1992. Hobart, Tasmania, November:

16. Mahidadia, A.J., C. Sammut , and P. Compton. *Building and Maintaining Causal Theories* in **AAAI Symposium on Knowledge Assimilation**. 1992. Stanford University, Spring, 1992.:

17. Menzies, T.J. *The Complexity of Model Review* in **Dx -93: The International Workshop on Principles on Model-Based Diagnosis**. 1993. Absersywath, Wales, UK:

18. Menzies, T.J., J. Black, J. Fleming, and M. Dean. *An Expert System for Raising Pigs* in **Th first Conference on Practical Applications of Prolog**. 1992. London, UK.:

19. Menzies, T.J. and B.R. Markey. *A Micro-Computer, Rule-Based Prolog Expert-System for Process Control in a Petrochemical Plant* in **Proceedings of the Third Australian Conference on Expert Systems, May 13-15**. 1987. Sydney, Australia:

20. O'Rourke, P., *Working Notes of the 1990 Spring Symposium on Automated Abduction*. 1990, University of California, Irvine, CA.:

21. Poole, D. *Hypo-Deductive Reasoning for Abduction, Default Reasoning, and Design* in **Working Notes of the 1990 Spring Symposium on Automated Abduction**. 1990. UC Irvine.

22. Poole, D., *A Methodology for Using a Default and Abductive Reasoning System* **International Journal of Intelligent Systems**, 1990. **5**: p. 521-548.

23. Preece, A.D. and R. Shinghal. *Verifying Knowledge Bases by Anomaly Detection: An Experience Report* in **ECAI '92**. 1992. Vienna:

24. Preston, P., G. Edwards, and P. Compton. *A 1600 rule expert system without knowledge engineers.* in **Second World Congress on Expert Systems**. 1993. Lisbon: Pergamon.

25. Ramsey, C.L., J.A. Reggia, D.S. Nau, and A. Ferrention, *A Comparative Analysis of Methods for Expert Systems* **Int. J. Man-Machine Studies**, 1986. **24**: p. 475-499.

26. Reggia, J.A. *Abductive Inference* in **Proceedings of the Expert Systems in Government Symposium**. 1985. Washington, D.C.:

27. Selman, B. and H.J. Levesque. *Abductive and Default Reasoning: a Computational Core* in **AAAI '90**. 1990.

28. Smythe, G.A., *Brain-hypothalmus, Pituitary and the Endocrine Pancreas*, in **The Endocrine Pancreas**, S. R., Editor. 1989, Raven Press: New York.

29. Suwa, M., A.C. Scott, and E.H. Shortliffe, *An Approach to Verifying Completeness and Consistency in a Rule-based Expert System*. 1982, Department of Computer Science, University of Stanford:

30. Zlatareva, N. *Distributed Verification and Automated Generation of Test Cases* in **IJCAI '93 workshop on Validation, Verification and Test of KBs**. 1993. Chambery, France: