

The Computational Complexity of Abduction

Tom Bylander, Dean Allemang,
Michael C. Tanner, and John R. Josephson
Laboratory for Artificial Intelligence Research
Department of Computer and Information Science
The Ohio State University
Columbus, Ohio
USA

June 20, 1990

to appear in *Artificial Intelligence*

Abstract

The problem of abduction can be characterized as finding the best explanation of a set of data. In this paper we focus on one type of abduction in which the best explanation is the most plausible combination of hypotheses that explains all the data. We then present several computational complexity results demonstrating that this type of abduction is intractable (NP-hard) in general. In particular, choosing between incompatible hypotheses, reasoning about cancellation effects among hypotheses, and satisfying the maximum plausibility requirement are major factors leading to intractability. We also identify a tractable, but restricted, class of abduction problems.

*Thanks to B. Chandrasekaran, Ashok Goel, Jack Smith, and Jon Sticklen for their comments on the numerous versions of this paper. The referees have also made a substantial contribution. Any remaining errors are our responsibility, of course. This research has been supported in part by the National Library of Medicine, grant LM-04298; the National Heart, Lung and Blood Institute, NIH Grant 1 R01 HL 38776-01; the National Science Foundation through a graduate fellowship; the Defense Advanced Research Projects Agency, RADC contract F30602-85-C-0010; the Air Force Office of Scientific Research, contract F49620-89-C-0110; and the National Science Foundation, grant IRI-8902142.

†This paper is a revised and integrated version of Allemang et al. [1] and Bylander et al. [2].

‡Dean Allemang's current address is Institut für Informatik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. Michael C. Tanner's current address is Computer Science Department, George Mason University, Fairfax, VA 22030, USA

1 Introduction

The problem of abduction can be characterized as finding the best explanation of a set of data [13]. Abduction applies to a wide variety of reasoning tasks [3]. For example, in medical diagnosis, the final diagnosis explains the signs and symptoms of the patient [23, 25]. In natural language understanding, the intended interpretation of a sentence explains why the sentence was said [12]. In scientific theory formation, the acceptance of a hypothesis is based on how well it explains the evidence [31].

What kinds of abduction problems can be solved efficiently? To answer this question, we must formalize the problem and then consider its computational complexity. However, it is not possible to prescribe a specific complexity threshold for all abduction problems. If the problem is “small,” then exponential time might be fast enough. If the problem is sufficiently large, then even $O(n^2)$ might be too slow. However, for the purposes of analysis, the traditional threshold of intractability, NP-hard, provides a rough measure of what problems are impractical [10]. Clearly, NP-hard problems will not scale up to larger, more complex domains.

Our approach is the following. First, we formally characterize abduction as a problem of finding the most plausible composite hypothesis that explains all the data. Then we consider several classes of problems of this type, the classes being differentiated by additional constraints on how hypotheses interact. We demonstrate that the time complexity of each class is polynomial (tractable) or NP-hard (intractable), relative to the complexity of computing the plausibility of hypotheses and the data explained by hypotheses.

Our results show that this type of abduction faces several obstacles. Choosing between incompatible hypotheses, reasoning about cancellation effects among hypotheses, and satisfying the maximum plausibility requirement are major factors making abduction intractable in general.

Some restricted classes of abduction problems are tractable. One kind of class is when some constraint guarantees a polynomial search space, e.g., the single-fault assumption (more generally, a limit on the size of composite hypotheses), or if all but a small number of hypotheses can be ruled out.¹ This kind of class trivializes complexity analysis because exhaustive search over the possible composite hypotheses becomes a tractable strategy.

However, we have discovered one class of abduction problems in which hypothesis assembly [13] can find the best explanation without exhaustive search. Informally, the constraints that define this class are: no incompatibility relationships, no cancellation interactions, the plausibilities of the individual hypotheses are all different from each other, and one explanation is qualitatively better than any other explanation. Unfortunately, it is intractable to determine whether the last condition holds. We consider one abduction system in which hypothesis assembly was applied, so as to examine the ramifications of these constraints in a real world situation.

The remainder of this paper is organized as follows. First, we provide a brief historical background to abduction. Then, we define our model of abduction problems and show how it applies to other theories of abduction. Next, we describe our complexity results, proofs

¹The latter constraint is not the same as “eliminating candidates” in de Kleer & Williams [6] or “inconsistency” in Reiter [26]. If a hypothesis is insufficient to explain all the observations, the hypothesis is not ruled out because it can still be in composite hypotheses.

of which are given in the appendix. Finally, we consider the relationship of these results to one abduction system.

2 Background

C. S. Peirce, who first described abductive inference [20], provided two intuitive characterizations: given an observation d and the knowledge that h causes d , it is an abduction to hypothesize that h occurred; and given a proposition q and the knowledge $p \rightarrow q$, it is an abduction to conclude p . In either case, an abduction is uncertain because something else might be the actual cause of d , or because the reasoning pattern is the classical fallacy of “affirming the consequent” and is formally invalid. Additional difficulties can exist because h might not always cause d , or because p might imply q only by default. In any case, we shall say that h *explains* d and p *explains* q , and we shall refer to h and p as *hypotheses* and d and q as *data*.

Pople pointed out the importance of abduction to AI [23], and he with Miller and Myers implemented one of the earliest abduction systems, INTERNIST-I, which performed medical diagnosis in the domain of internal medicine [16, 24]. This program contained an explicit list of diseases and symptoms, explicit causal links between the diseases and the symptoms, and probabilistic information associated with the links. INTERNIST-I used a form of hill climbing—once a disease outscored its competitors by a certain threshold, it was permanently selected as part of the final diagnosis. Hypothesis assembly [13] is a generalization of this technique. Below, we describe a restricted class of problems for which hypothesis assembly can efficiently find the best explanation.

Based on similar explicit representations, Pearl [19] and Peng & Reggia [21] find the most probable composite hypothesis that explains all the data, a task that is known to be intractable in general [4]. Below we describe additional constraints under which this task remains intractable.

In contrast to maintaining explicit links between hypotheses and data, Davis & Hamscher’s model-based diagnosis [5] determines at run-time what data need to be explained and what hypotheses can explain the data. Much of this work, such as de Kleer & Williams [6] and Reiter [26], place an emphasis on generating all “minimal” composite hypotheses that explain all the data. However, there can be an exponential number of such hypotheses. Current research is investigating how to focus the reasoning on the most relevant composite hypotheses [7, 8, 30]. However, we show below that it is intractable in general to find a composite hypothesis that explains all the data, and that even if it is easy to find explanations, generating all the relevant composite hypotheses is still intractable.

Whatever the technique or formulation, certain fundamentals of the abduction task do not change. In particular, our analysis shows how computational complexity arises from constraints on the explanatory relationship from hypotheses to data and on plausibility ordering among hypotheses. These constraints do not depend on the style of the representation or reasoning method (causal vs. logical, probabilistic vs. default, explicit vs. model-based, ATMS or not, etc.). In other words, certain kinds of abduction problems are hard no matter what representation or reasoning method is chosen.

3 Notation, Definitions, and Assumptions

We use the following notational conventions and definitions. d stands for a datum, e.g., a symptom. D stands for a set of data. h stands for an individual hypothesis, e.g., a hypothesized disease. H stands for a set of individual hypotheses. H can be treated as a composite hypothesis, i.e., each $h \in H$ is hypothesized to be present, and each $h \notin H$ is hypothesized to be absent or irrelevant.

3.1 Model of Abduction

An *abduction problem* is a tuple $\langle D_{all}, H_{all}, e, pl \rangle$, where:

D_{all} is a finite set of all the data to be explained,

H_{all} is a finite set of all the individual hypotheses,

e is a map from subsets of H_{all} to subsets of D_{all} (H explains $e(H)$), and

pl is a map from subsets of H_{all} to a partially ordered set (H has plausibility $pl(H)$).

For the purpose of this definition and the results below, it does not matter whether $pl(H)$ is a probability, a measure of belief, a fuzzy value, a degree of fit, or a symbolic likelihood. The only requirement is that the range of pl is partially ordered.

H is *complete* if $e(H) = D_{all}$. That is, H explains all the data.

H is *parsimonious* if $\nexists H' \subset H$ ($e(H) \subseteq e(H')$). That is, no proper subset of H explains all the data that H does.

H is an *explanation* if it is complete and parsimonious. That is, H explains all the data and has no explanatorily superfluous elements. Note that an explanation exists if and only if a complete composite hypothesis exists.²

H is a *best explanation* if it is an explanation, and if there is no explanation H' such that $pl(H') > pl(H)$. That is, no other explanation is more plausible than H . It is just “a best” because pl might not impose a total ordering over composite hypotheses (e.g., because of probability intervals or qualitative likelihoods). Consequently, several composite hypotheses might satisfy this definition.

3.2 Relation to Other Work

These definitions are intended to formalize the notion of best explanation in Josephson et al. [13]. However, our definitions are not limited to that paper. We consider in detail here how Reiter’s theory of diagnosis [26] and Pearl’s theory of belief revision [19] can be mapped to our model of abduction.

²Composite hypotheses that do not explain all the data can still be considered explanations, albeit partial. Nevertheless, because explaining all the data is a goal of the abduction problems we are considering, for convenience, this goal is incorporated into the definition of “explanation.”

3.2.1 Reiter’s Theory of Diagnosis

Reiter defines a diagnosis problem as a tuple $\langle \text{SD}, \text{COMPONENTS}, \text{OBS} \rangle$, in which SD and OBS are finite sets of first-order sentences comprising the system description and observations, respectively; COMPONENTS is a finite set of constants; and AB is a distinguished unary predicate, interpreted as abnormal. A diagnosis is defined to be a minimal set $\Delta \subseteq \text{COMPONENTS}$ such that:

$$\text{SD} \cup \text{OBS} \cup \{\text{AB}(c) \mid c \in \Delta\} \cup \{\neg\text{AB}(c) \mid c \in \text{COMPONENTS} \setminus \Delta\}$$

is consistent. “Minimal set” means that no subset of Δ satisfies the same condition.

Each subset of COMPONENTS can be treated as a composite hypothesis, i.e., a conjecture that certain components are abnormal, and that all other components are normal. A diagnosis problem can then be mapped into an abduction problem as follows:

$$\begin{aligned} H_{all} &= \text{COMPONENTS} \\ D_{all} &= \text{OBS} \\ e(H) &= \text{a maximal set } D \subseteq D_{all} \text{ such that}^3 \\ &\quad \text{SD} \cup D \cup \{\text{AB}(h) \mid h \in H\} \cup \{\neg\text{AB}(h) \mid h \in H_{all} \setminus H\} \\ &\quad \text{is consistent.} \end{aligned}$$

A solution for the diagnosis problem then corresponds to an explanation for the abduction problem, and vice versa. Reiter does not define any criteria for ranking diagnoses, so there is nothing to map to pl .

3.2.2 Pearl’s Theory of Belief Revision

A Bayesian belief network [18] is a directed acyclic graph whose nodes \mathbf{W} are propositional variables. The probabilistic dependencies between the variables are described by specifying $P(x|s)$ for each value assignment x to a variable $X \in \mathbf{W}$ and each value assignment s to X ’s parents \mathbf{S} .⁴ The intention is that “the arcs signify the existence of direct causal influences between the linked propositions, and the strengths of these influences are quantified by the conditional probabilities of each variable given the state of its parents” [19, p. 175].

For a particular belief revision problem [19], some subset \mathbf{V} of the variables \mathbf{W} are initialized with specific values. Let v be the value assignment to \mathbf{V} . The solution to the problem is the most probable value assignment w^* to all the variables \mathbf{W} , i.e., $P(w^*|v)$ is greater than or equal to $P(w|v)$ for any other value assignment w to the variables \mathbf{W} . w^* is called the most probable explanation (MPE).

v can be mapped to the set of data to be explained, i.e., a value assignment x to a variable $X \in \mathbf{V}$ is a datum. v can be explained by appropriate value assignments to the other variables $\mathbf{W} \setminus \mathbf{V}$. Treating value assignments of *true* as individual hypotheses, a belief revision problem can be mapped to an abduction problem as follows:

³There might be more than one maximal subset of observations that satisfies these conditions. If so, then $e(H)$ selects some preferred subset.

⁴For belief networks, we use a (boldface) lower case letter to stand for a (set of) value assignment(s) to a (set of) variable(s), which is denoted by a (boldface) upper case letter.

$$\begin{aligned}
D_{all} &= v \\
H_{all} &= W \setminus V \\
e(H) &= \text{a maximal set } D \subseteq D_{all} \text{ such that} \\
&\quad P(H = true \wedge H_{all} \setminus H = false | D) > 0 \\
pl(H) &= P(H = true \wedge H_{all} \setminus H = false | e(H))
\end{aligned}$$

The MPE corresponds to a complete composite hypothesis. If the MPE is also parsimonious, then it corresponds to the best explanation.⁵ However, the MPE might assign *true* to more variables than necessary for explanatory purposes. In the context of other value assignments, $X = true$ might be more likely than $X = false$ even if $X = true$ is superfluous under the above mapping [21].

This lack of correspondence between the MPE and the best explanation can be rectified by creating, for each $X \in W \setminus V$, a dummy variable X' and a dummy value assignment that can be “caused” only if $X \neq X'$. With this modification, the MPE corresponds to the best explanation.

Another way of rectifying the situation is to simply ignore the parsimony constraint. With this in mind, we shall use the mapping given above.

3.2.3 Other Theories of Abduction

These reductions from problems in Reiter’s and Pearl’s theories to abduction problems provide strong evidence that our model of abduction is general enough to accommodate any theory of abduction, e.g., [6, 15, 21, 22]. This is because our model leaves e and pl virtually unconstrained. We exploit this freedom below by defining and analyzing natural constraints on e and pl without considering the representations—logical, causal, or probabilistic—underlying the computation of e and pl . To make the analysis complete, we also show how some of these constraints can be reduced to problems in Reiter’s and Pearl’s theories.

3.3 Tractability Assumptions

In our complexity analysis, we assume that e and pl are tractable. We also assume that e and pl can be represented reasonably, in particular, that the size of their internal representations is polynomial in $|D_{all}| + |H_{all}|$.

Clearly, the tractability of these functions is central to abduction, since it is difficult to find plausible hypotheses explaining the data if it is difficult to compute e and pl . This should not be taken to imply that the tractability of these functions can be taken for granted. For example, it can be intractable to determine explanatory coverage of a composite hypothesis [26] and to calculate the probability that an individual hypothesis is present, ignoring other hypotheses [4]. We make these assumptions to simplify our analysis of abduction problems. To reflect the complexity of these functions in our tractability results, we denote the time complexity of e and pl with respect to the size of an abduction problem as \mathcal{C}_e and \mathcal{C}_{pl} , respectively, e.g., $n\mathcal{C}_e$ indicates n calls to e .

⁵One difficulty with the more “natural” mapping $pl(H) = P(H = true | v)$ is that even if the MPE is parsimonious, it might not be the best explanation.

For convenience, we assume the existence and the tractability of a function that determines which individual hypotheses can contribute to explaining a datum. Although it is not a true inverse, we refer to this function as e^{-1} , formally defined as:

$$e^{-1}(d) = \{h \mid \exists H \subset H_{all} (d \notin e(H) \wedge d \in e(H \cup \{h\}))\}$$

Note that $h \in e^{-1}(d)$ does not imply $d \in e(h)$.

The key factors, then, that we consider in the complexity of finding a best explanation are properties of e and pl that allow or prevent tractable computation given that e , e^{-1} , and pl can be computed “easily.” That is, given a particular class of abduction problems, how much of the space of composite hypotheses must be explicitly searched to find a best explanation? As demonstrated below, intractability is the usual result in classes of problems that involve significant interaction among the elements of composite hypotheses.

3.4 Simplifications

We should note that these definitions and assumptions simplify several aspects of abduction. For example, we define composite hypotheses as simple combinations of individual hypotheses. In reality, the relationships among the parts of an abductive answer and the data being explained can be much more complex, both logically and causally.

Another simplification is that *domains* are not defined. One way to do this would be to specify what data are possible (D_{poss}) and general functions for computing explanatory coverage and plausibilities based on the data (e_{gen} and pl_{gen}). Then for a specific abduction problem, the following constraints would hold: $D_{all} \subseteq D_{poss}$, $e(H) = e_{gen}(H, D_{all})$, and $pl(H) = pl_{gen}(H, D_{all})$ (cf. Allemang et al. [1]).

The definitions of abduction problems or domains do not mention the data that do not have to be explained, even though they could be important for determining e and pl . For example, the age of a patient does not have to be explained, but can influence the plausibility of a disease. We shall assume that e and pl implicitly take into account data that do not have to be explained, e.g., in the definition of domains above, these data can be an additional argument to e_{gen} and pl_{gen} .

Despite these simplifications, our analysis provides powerful insights concerning the computational complexity of abduction.

3.5 An Example

We shall use the following example to facilitate our discussion:

$$\begin{aligned} H_{all} &= \{h_1, h_2, h_3, h_4, h_5\} \\ D_{all} &= \{d_1, d_2, d_3, d_4\} \\ e(h_1) &= \{d_1\} & pl(h_1) &= \text{superior} \\ e(h_2) &= \{d_1, d_2\} & pl(h_2) &= \text{excellent} \\ e(h_3) &= \{d_2, d_3\} & pl(h_3) &= \text{good} \\ e(h_4) &= \{d_2, d_4\} & pl(h_4) &= \text{fair} \\ e(h_5) &= \{d_3, d_4\} & pl(h_5) &= \text{poor} \end{aligned}$$

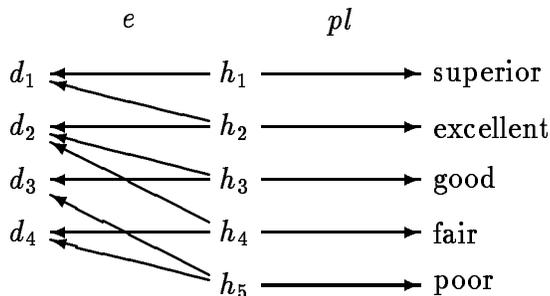


Figure 1: Example of an Abduction Problem

Figure 1 is a pictorial representation of the example. The values of pl should simply be interpreted as indicating relative order of plausibility. If $e(H)$ is the union of $e(h)$ for $h \in H$, then $\{h_2, h_3, h_5\}$ is complete, but not parsimonious since h_3 is superfluous. $\{h_2, h_3\}$ is parsimonious, but not complete since it does not explain d_4 . Based on the plausibility ordering criterion defined in Section 5, $\{h_1, h_3, h_4\}$ and $\{h_2, h_5\}$ would be considered the best explanations.

Using these definitions, assumptions, and example, we first discuss how properties of e affect the tractability of finding best explanations, and then consider properties of pl .

4 Complexity of Finding Explanations

4.1 Independent Abduction Problems

In the simplest problems, an individual hypothesis explains a specific set of data regardless of what other individual hypotheses are being considered. This constraint is assumed by INTERNIST-I [16], Reggia’s set covering algorithm [25], Peng & Reggia’s parsimonious covering theory [21], Pearl’s belief revision theory if interactions are restricted to noisy-OR (an effect can occur only if one or more of its causes are present) [19], and Eshelman’s cover-and-differentiate method [9]. The use of conflict sets [6, 26] also appears to make this assumption—each conflict set corresponds to a datum to be explained, and the elements of the conflict set correspond to the hypotheses that can independently explain the datum.

Formally, an abduction problem is *independent* if:

$$\forall H \subseteq H_{all} (e(H) = \bigcup_{h \in H} e(h))$$

That is, a composite hypothesis explains a datum if and only if one of its elements explains the datum. This constraint makes explanatory coverage equivalent to set covering [25]. Assuming independence, the explanations in our example (refer to Figure 1) are: $\{h_1, h_3, h_4\}$, $\{h_1, h_3, h_5\}$, $\{h_1, h_4, h_5\}$, $\{h_2, h_3, h_4\}$, and $\{h_2, h_5\}$.

One way to find a best explanation would be to generate all explanations and then sort them by plausibility. However, it is well-known that there can be an exponential number of explanations. It is not surprising then that determining the number of explanations is hard.

*W stands for the working composite hypothesis.
Nil is returned if no explanation exists.*

*Determine whether an explanation exists.
If $e(H_{all}) \neq D_{all}$ then
Return nil*

*Find an explanation.
 $W \leftarrow H_{all}$
For each $h \in H_{all}$
If $e(W \setminus \{h\}) = D_{all}$ then
 $W \leftarrow W \setminus \{h\}$
Return W*

Algorithm 1: Finding an Explanation in Independent and Monotonic Abduction Problems

Theorem 4.1 *For the class of independent abduction problems, it is #P-complete to determine the number of explanations.*

That is, determining the number of explanations for an independent abduction problem is just as hard as determining the number of solutions to an NP-complete problem.⁶

The definition of best explanation, however, does not require that all explanations be explicitly enumerated. For example, if h is the most plausible individual hypothesis, h explains D_{all} , and $H \subset H'$ implies $pl(H) > pl(H')$, then h can be declared to be the best explanation without further search. In general, the task of finding a best explanation can be divided into two subtasks: (1) find one explanation, and (2) repeatedly find better explanations until a best one is found. In the remainder of this section then, we shall consider the complexity of generating one or more explanations. The following section discusses the complexity of finding better explanations.

For independent abduction problems, it is tractable to find an explanation. Let $n = |D_{all}| + |H_{all}|$.

Theorem 4.2 *For the class of independent abduction problems, there is an $O(nC_e + n^2)$ algorithm for finding an explanation, if one exists.*

Algorithm 1 performs this task within this order of complexity. The appendix gives a detailed explanation of this algorithm, but we note several aspects of its operation here.

It is easy to check whether an explanation exists. If $\bigcup_{h \in H_{all}} e(h) \neq D_{all}$, then a union over any subset of H_{all} will not equal D_{all} either.

The loop makes one pass through the individual hypotheses. It examines each individual hypothesis in turn and removes it if no explanatory coverage is lost. Only one pass is

⁶Also, it is #P-complete to determine the number of complete composite hypotheses. The definition of #P-complete comes from Valiant [32].

Detailed proofs of Theorem 4.1 and other theorems are given in the appendix.

necessary because if the result W had a superfluous element h , then h would have been superfluous for any superset of W , and thus, would have been removed by the body of the loop.

If the e^{-1} function is available (see Section 3.3 for the definition of e^{-1}), then the working hypothesis W , instead of being initialized to H_{all} , can be initialized to include only one element from $e^{-1}(d)$ for each $d \in D_{all}$. This modification has an advantage if e^{-1} is easy to compute and the working hypothesis remains “small.”

4.2 Monotonic Abduction Problems

We now consider a more general kind of problem, in which a composite hypothesis can explain additional data that are not explained by any of its elements. For example, suppose the two inputs to an AND-gate are supposed to be 0, and so the output is supposed to be 0, but the observed output of the AND-gate is 1. If the inputs are produced by components A and B , then hypothesizing a single fault in A or B is insufficient to account for the datum, but faults in both A and B are sufficient.

This sort of interaction can also occur if two individual hypotheses have an additive interaction. For example, each of the two hypotheses can explain a small value of some measurement, but together can explain a larger measurement. In this latter case, if h only partially explains d , then $d \notin e(h)$. Note though that if adding h to a composite hypothesis can result in completely explaining d , then $h \in e^{-1}(d)$.

Formally, an abduction problem is *monotonic* [1] if:

$$\forall H, H' \subseteq H_{all} (H \subseteq H' \rightarrow e(H) \subseteq e(H'))$$

That is, a composite hypothesis does not “lose” any data explained by any of its subsets and might explain additional data. All independent abduction problems are monotonic, but a monotonic abduction problem is not necessarily independent. If, in Figure 1, $\{h_2, h_3\}$ also explained d_4 , then $\{h_2, h_3\}$ would also be an explanation and $\{h_2, h_3, h_4\}$ would not be. Monotonic abduction problems from the literature include Josephson’s hypothesis assembly technique [1] and Pearl’s belief revision theory if interactions are restricted to noisy-OR and noisy-AND [19].

Because the class of monotonic abduction problems includes the independent class, it is also hard to determine the number of explanations. In addition, we have shown that it is hard to enumerate a polynomial number of explanations.

Theorem 4.3 *For the class of monotonic abduction problems, given a set of explanations, it is NP-complete to determine whether an additional explanation exists.*

We have proven this result by a reduction from the class of independent incompatibility abduction problems, which is described below. The idea of the reduction is that the addition of an individual hypothesis to a composite hypothesis can explain the rest of the data, make nearly all the elements of the composite hypothesis superfluous, and result in a previously generated explanation. It turns out to be difficult to generate an additional explanation while avoiding this kind of interaction. Whether a similar result holds for independent abduction problems is an open question.

Although the class of monotonic problems is a superset of the class of independent problems, it is just as efficient to find an explanation. Again, let $n = |D_{all}| + |H_{all}|$.

Theorem 4.4 *For the class of monotonic abduction problems, there is an $O(nC_e + n^2)$ algorithm for finding an explanation, if one exists.*

Algorithm 1 performs this task within this order of complexity. Because of the monotonicity constraint, H_{all} must explain as much or more data than any other composite hypothesis. The loop in Algorithm 1 works for the same reasons as for independent abduction problems. Also, it is possible to use e^{-1} to initialize W , though one must be careful because more than one element from $e^{-1}(d)$ might be needed to explain d .

4.3 Incompatibility Abduction Problems

Implicit in the formal model so far is the assumption that any collection of individual hypotheses is possible. However, most domains have restrictions that invalidate this assumption. For example, a faulty digital switch cannot simultaneously be stuck-at-1 and stuck-at-0. More generally, the negation of a hypothesis can also be considered a hypothesis.

This kind of problem is neither independent nor monotonic because any composite hypothesis that contains a pair of mutually exclusive hypotheses cannot be an acceptable hypothesis, while a subset that excludes at least one hypothesis from each pair is acceptable. We call this kind of problem an *incompatibility abduction problem*.

Formally, an incompatibility abduction problem is a tuple $\langle D_{all}, H_{all}, e, pl, \mathcal{I} \rangle$, where D_{all} , H_{all} , e , and pl are the same as before and \mathcal{I} is a set of two-element subsets of H_{all} , indicating pairs of hypotheses that are incompatible with each other.⁷ For an incompatibility problem:

$$\forall H \subseteq H_{all} ((\exists I \in \mathcal{I} (I \subseteq H)) \rightarrow e(H) = \emptyset)$$

By this formal trick, a composite hypothesis containing incompatible hypotheses explains nothing, preventing such a composite from being complete (except for trivial cases) or a best explanation.

An *independent incompatibility abduction problem* satisfies the formula:

$$\forall H \subseteq H_{all} ((\nexists I \in \mathcal{I} (I \subseteq H)) \rightarrow e(H) = \bigcup_{h \in H} e(h))$$

That is, except for incompatibilities, the problem is independent. In Figure 1, if $\mathcal{I} = \{\{h_1, h_2\}, \{h_2, h_3\}, \{h_3, h_4\}, \{h_4, h_5\}\}$, then only $\{h_1, h_3, h_5\}$ and $\{h_2, h_5\}$ would be explanations.

Incompatibility abduction problems are more complex than monotonic or independent abduction problems:

Theorem 4.5 *For the class of independent incompatibility abduction problems, it is NP-complete to determine whether an explanation exists.*

⁷Incompatible pairs are the most natural case, e.g., one hypothesis of the pair is the negation of the other. n mutually exclusive hypotheses can be represented as $n(n-1)/2$ incompatible pairs. Incompatible triplets (any two of the three, but not all three) and so on are conceivable, but allowing these possibilities in the formal definition do not affect the complexity results.

We have proven this result by reduction from 3SAT [10], which is satisfiability of boolean expressions in conjunctive normal form, with no more than three literals in any conjunct. Informally, the reduction works as follows. Each 3SAT literal and its negation corresponds to an incompatible pair of hypotheses. Each conjunct of the boolean expression corresponds to a datum to be explained. Satisfying a conjunct corresponds to a hypothesis explaining a datum. Clearly then, a complete composite hypothesis exists iff the boolean expression is satisfiable. Furthermore, a complete composite hypothesis exists iff an explanation exists. Our proof shows that only $O(|H_{all}|)$ incompatible pairs are needed to give rise to intractability.

The underlying difficulty is that the choice between a pair of incompatible hypotheses cannot be made locally, but is dependent on the choices from all other incompatible pairs. It is interesting to note the parsimony constraint plays no role in this result. Just finding a complete composite hypothesis is hard in incompatibility abduction problems.

It follows that:

Corollary 4.6 *For the class of independent incompatibility abduction problems, it is NP-hard to find a best explanation.*

The class of incompatibility abduction problems can be reduced to both Reiter’s theory of diagnosis [26] and Pearl’s theory of belief revision [19].

Theorem 4.7 *For the class of diagnosis problems, relative to the complexity of determining whether a composite hypothesis is consistent with SDUOBS, it is NP-complete to determine whether a diagnosis exists.*

For this theorem, a composite hypothesis is a conjecture that certain components are abnormal, and that all other components are normal.

It is easy to translate the explanatory interactions of an independent incompatibility abduction problem into first-order sentences. For example, $e^{-1}(d) = H$ can be translated to $\text{MANIFEST}(d) \rightarrow \bigvee_{h \in H} \text{AB}(h)$. $\{h, h'\} \in \mathcal{I}$ can be translated to $\text{AB}(h) \rightarrow \neg \text{AB}(h')$. It is interesting that this problem is hard even if it is easy to determine the consistency of a composite hypothesis.

Theorem 4.8 *For the class of belief revision problems, it is NP-complete to determine whether there is a value assignment w to the variables W such that $P(w|v) > 0$.*

This theorem directly follows from Cooper’s result that it is NP-complete to determine whether $P(X = \text{true}) > 0$ for a given variable X within a belief network [4]. Also, a reduction from incompatibility abduction problems can be done as follows. Map each $h \in H_{all}$ to a “hypothesis” variable. Map each $d \in D_{all}$ to a “data” variable that can be true only if one or more of the hypothesis variables corresponding to $e^{-1}(d)$ are true (e.g., noisy-OR interaction). Map each incompatible pair into a data variable that can be true only if at most one, but not both, of the two corresponding hypothesis variables is true (e.g., NAND). Initializing all the data variables to *true* sets up the problem.

4.4 Cancellation Abduction Problems

Another interaction not allowed in independent or monotonic abduction problems is cancellation, i.e., when one element of a composite hypothesis “cancels” a datum that another element would otherwise explain. Cancellation can occur when one hypothesis can have a subtractive effect on another. This is common in medicine, e.g., in the domain of acid-base disorders, one disease might explain an increased blood pH, and another might explain a decreased pH, but together the result might be a normal pH [17]. Different faults in different components can result in cancellation, e.g., a stuck-at-1 input into an AND-gate might account for an output of 1, but not if the other input is stuck-at-0. Cancellation commonly occurs in the physical world. Newton’s second law implies that forces can cancel each other. Cancellation in the form of feedback control is intentionally designed into devices.

Formally, we define a *cancellation abduction problem* as a tuple $\langle D_{all}, H_{all}, e, pl, e_+, e_- \rangle$. e_+ is a map from H_{all} to subsets of D_{all} indicating what data each hypothesis “produces.” e_- is another map from H_{all} to subsets of D_{all} indicating what data each hypothesis “consumes.” $d \in e(H)$ iff the number of hypotheses in H that produce d outnumber the hypotheses that consume d . That is:

$$d \in e(H) \leftrightarrow |\{h \mid h \in H \wedge d \in e_+(h)\}| > |\{h \mid h \in H \wedge d \in e_-(h)\}|$$

In Figure 1, if we let $e_+ = e$ for individual hypotheses and if $e_-(h_1) = \{d_3\}$, $e_-(h_2) = \{d_4\}$, and $e_-(h_3) = e_-(h_4) = e_-(h_5) = \emptyset$, then the only explanations would be $\{h_1, h_3, h_5\}$ and $\{h_2, h_4, h_5\}$.

Admittedly, this is a simplified model of cancellation effects, in the sense that it captures only one kind of cancellation interaction. Nevertheless, it is sufficient to derive intractability:

Theorem 4.9 *For the class of cancellation abduction problems, it is NP-complete to determine whether an explanation exists.*

We have proven this by reduction from finding explanations in incompatibility abduction problems. Informally, the idea of the reduction is based on the following. Suppose that a datum has two potential “producers” and two potential “consumers.” Now any composite hypothesis that contains both consumers cannot explain the datum. In effect, the two consumers are incompatible. Our reduction ensures that each incompatible pair in the incompatibility abduction problem is appropriately mapped to such a situation in the corresponding cancellation abduction problem. Only $O(|H_{all}|)$ “cancellations” are needed for this result, where $\sum_{h \in H_{all}} |e_-(h)|$ gives the number of cancellations.

It follows that:

Corollary 4.10 *For the class of cancellation abduction problems, it is NP-hard to find a best explanation.*

One aspect of cancellation abduction problems is more complex than incompatibility abduction problems. In an independent incompatibility abduction problem, if a complete composite hypothesis is found, then it is easy to find a parsimonious subset. However, this is not true for cancellation abduction problems.

Table 1: Computational Complexity of Finding Explanations

class of problems	condition to achieve		
	finding all explanations	finding an explanation	finding a best explanation
independent	NP	P	?
monotonic	NP	P	?
incompatibility	NP	NP	NP
cancellation	NP	NP	NP

P = known polynomial algorithm NP = NP-hard

Theorem 4.11 *For the class of cancellation abduction problems, it is coNP-complete to determine whether a complete composite hypothesis is parsimonious.*

That is, it is NP-complete to determine whether a complete composite hypothesis is not parsimonious. The idea of our reduction is the following. If a datum has three “producers” and two “consumers,” we can ensure that the datum is explained by including all three producers in the composite hypothesis. However, there might be a more parsimonious composite hypothesis in which some of the producers are omitted, but finding such a composite hypothesis means that one or both consumers must be omitted as well, making them effectively incompatible.

Table 1 summarizes the results of this section. The “?” indicates that we not have yet described the complexity of finding a best explanation in independent and monotonic abduction problems.

5 Complexity of Plausibility

To analyze the complexity of finding a best explanation, we need to define how to compare the plausibilities of explanations. We consider one plausibility criterion based on comparing the plausibilities of the elements of the explanations. Other plausibility criteria are considered in Bylander et al. [2], but they are less relevant to other theories of abduction.

5.1 The Best-Small Plausibility Criterion

Everything else being equal, smaller explanations are preferable to larger ones, and more plausible individual hypotheses are preferable to less plausible ones. Thus, in the absence of other information, it is reasonable to compare the plausibility of explanations based on their sizes and the relative plausibilities of their elements. When a conflict occurs, e.g., one explanation is smaller, but has less plausible elements, no ordering can be imposed without additional information.

The *best-small* plausibility criterion formally characterizes these considerations as follows:

$$\begin{aligned}
pl(H) > pl(H') &\leftrightarrow \\
\exists m: H \rightarrow H' &(m \text{ is 1-1} \wedge \\
&\forall h \in H (pl(h) \geq pl(m(h))) \wedge \\
&(|H| = |H'| \rightarrow \exists h \in H (pl(h) > pl(m(h))))))
\end{aligned}$$

That is, to be more plausible according to best-small, the elements of H need to be matched to the elements of H' so that the elements of H are at least as plausible as their matches in H' . If H and H' are the same size, then in addition some element in H must be more plausible than its match in H' . Note that if H is larger than H' , then $pl(H) \not> pl(H')$. In Figure 1, $\{h_1, h_3, h_4\}$ and $\{h_2, h_5\}$ would be the best explanations.

We have demonstrated that it is intractable to find best explanations using best-small.

Theorem 5.1 *For the class of independent abduction problems using the best-small plausibility criterion, it is NP-hard to find a best explanation.*

The simplest proof of this theorem involves a reduction from minimum cover [10]. If each individual hypothesis is given the same plausibility, then the smallest explanations (the covers with the smallest sizes) are the best explanations. A more general proof is a reduction from a special class of independent incompatibility abduction problems in which each individual hypothesis is in exactly one incompatible pair. In this reduction, each incompatible pair is mapped into two equally plausible hypotheses, at least one of which must be chosen. If the incompatibility abduction problem has any explanations, they turn out to be best explanations in the best-small problem.

We conjecture that it is possible to reduce from finding a best explanation using best-small to finding a best explanation using any “theory of belief” in which composite hypotheses that are smaller or have more plausible elements can have higher belief values. Of course, standard probability theory is an example of such a theory, as are all its main competitors. This conjecture is supported by the following theorem.

Theorem 5.2 *For the class of belief revision problems restricted to OR interactions, it is NP-hard to find the MPE.*

The restriction to OR interactions means that each effect can be true only if one or more of its parents are true. This restriction makes it easy to find a value assignment w such that $P(w|v) > 0$. Although this theorem could be demonstrated by adapting the proof for Theorem 5.1, it is useful to show that the best-small plausibility criterion has a correlate in probabilistic reasoning.

The reduction from independent abduction problems using best-small works as follows. Each $h \in H_{all}$ is mapped to a “hypothesis” variable. Each $d \in D_{all}$ is mapped to a “data” variable that is true if and only if one or more of the hypothesis variables corresponding to $e^{-1}(d)$ are true, i.e., an OR interaction. The *a priori* probabilities of the hypothesis variables being true must be between 0 and .5, and are ordered according to the plausibilities in the abduction problem. Initializing all the data variables to *true* sets up the problem. The MPE for this belief revision problem corresponds to a best explanation for the best-small problem. Because finding a best explanation is NP-hard, finding the MPE must be NP-hard even for belief networks that only contain OR interactions.

5.2 Ordered Abduction Problems

Our proofs of Theorem 5.1 depend on the fact that some individual hypotheses have similar plausibilities to other hypotheses. It turns out that finding a best explanation using best-small is tractable if the plausibilities of individual hypotheses are all different from each other and if their plausibilities are totally ordered.

Formally, an abduction problem is *ordered* if:

$$\forall h, h' \in H_{all} (h \neq h' \rightarrow (pl(h) < pl(h') \vee pl(h) > pl(h')))$$

Again, let $n = |D_{all}| + |H_{all}|$.

Theorem 5.3 *For the class of ordered monotonic abduction problems using the best-small plausibility criterion, there is an $O(nC_e + nC_{pl} + n^2)$ algorithm for finding a best explanation.*

Algorithm 2 performs this task within this order of complexity. It is same as Algorithm 1 except that the loop considers the individual hypotheses from least to most plausible. The explanation that Algorithm 2 finds is a best explanation because no other explanation can have more plausible individual hypotheses; the algorithm always chooses the least plausible individual hypotheses to remove. Of course, Algorithm 2 also finds a best explanation for ordered independent abduction problems.

As with Algorithm 1, it is possible to use e^{-1} advantageously. The working hypothesis W can be initialized to include the most plausible individual hypotheses from each $e^{-1}(d)$, i.e., because of monotonic interactions, sufficient hypotheses from $e^{-1}(d)$ must be chosen so that d is explained.

Algorithm 2 is an adaptation of the hypothesis assembly algorithm described in Josephson et al. [13], and is a serial version of the parallel parsimony algorithm described in Goel [11]. In Figure 1 assuming the independence constraint, this algorithm would find $\{h_1, h_3, h_4\}$, which is one of the two best explanations.

As in our example, there might be more than one explanation because best-small in general imposes a partial ordering on the plausibilities of composite hypotheses. Suppose that an ordered monotonic abduction problem had only one best explanation according to best-small. Because Algorithm 2 is guaranteed to find a best explanation, then it will find the one best explanation.

Corollary 5.4 *For the class of ordered monotonic abduction problems using the best-small plausibility criterion, if there is exactly one best explanation, then there is an $O(nC_e + nC_{pl} + n^2)$ algorithm for finding the best explanation.*

This can be informally restated as: *In a well-behaved abduction problem, if it is known that some explanation is clearly the best explanation, then it is tractable to find it.* Unfortunately, it is difficult to determine if some explanation is clearly the best explanation.

Theorem 5.5 *For the class of ordered independent abduction problems using the best-small plausibility criterion, given a best explanation, it is NP-complete to determine whether there is another best explanation.*

*W stands for the working composite hypothesis.
Nil is returned if no explanation exists.*

*Determine whether an explanation exists.
If $e(H_{all}) \neq D_{all}$ then
Return nil*

*Find a best explanation.
 $W \leftarrow H_{all}$
For each $h \in H_{all}$ from least to most plausible
If $e(W \setminus \{h\}) = D_{all}$ then
 $W \leftarrow W \setminus \{h\}$
Return W*

Algorithm 2: Finding a Best Explanation in Ordered Independent and Monotonic Abduction Problems Using the Best-Small Plausibility Criterion

We have proved this by a reduction from the special class of independent incompatibility abduction problems in which each individual hypothesis is in exactly one incompatible pair. Assuming n incompatible pairs, the best-small problem is set up so that one hypothesis out of each pair must be chosen, and so that extra hypotheses plus the most plausible element of each pair is a best explanation of size $n + 2$. In our reduction, any other best-small best explanation in this reduction must be of size $n + 1$ and include an explanation for the incompatibility problem. Thus, even for ordered independent abduction problems, it is intractable to find all the best explanations, or even enumerate some number of them.

As a consequence, it does not become tractable to find the MPE for ordered abduction problems. The proof for the previous theorem can be easily adapted so that any explanation of size $n + 1$ will be more probable than any explanation of size $n + 2$.

From these theorems, we can describe what kinds of mistakes will be made by Algorithm 2. While the explanation this algorithm finds will match up qualitatively to any other explanation, there might be other “qualitatively best” explanations, which might be judged better based on more precise plausibility information.

Table 2 summarizes the results of this and the previous section.

6 A Real-World Application of Abduction—Red Blood Antibody Identification

6.1 Description of the Domain

The RED expert system performs in the domain of blood bank antibody analysis [28]. One of the jobs done by a blood-bank technologist is to identify antibodies in a patient’s serum that can react to antigens that might appear on red blood cells. This is typically

Table 2: Computational Complexity of Finding Best Explanations Using the Best-Small Plausibility Criterion

class of problems	condition to achieve	
	finding a best explanation	finding more than one best explanation
ordered independent/monotonic	P	NP
unordered independent/monotonic	NP	NP
incompatibility	NP	NP
cancellation	NP	NP

P = known polynomial algorithm NP = NP-hard

done by combining, under different test conditions, samples of patient serum with samples of red blood cells known to express certain antigens. Some of these combinations might show reactions. The presence of certain antibodies in the patient serum accounts for certain reactions. The reactions are additive in the sense that if the presence of one antibody explains one reaction, and presence of another antibody explains another, then the presence of both antibodies explains both reactions. If each antibody can account for a weak result in some reaction, then the presence of both can account for a stronger result in that reaction. Also, some pairs of antibodies cannot occur together. RED’s task is to decide which antibodies are present, given a certain reaction pattern. RED takes into account about 30 of the most common antibodies.

6.2 Relationship to Classes of Abduction Problems

We now examine how this task can be categorized within the classes of abduction problems discussed in this paper.

Independent. The additive nature of the reactions means that for separate reactions and compatible hypotheses, the independence constraint is met. However, since independent abduction problems do not allow for parts of data to be explained, they cannot describe additivity of reaction strengths.

Monotonic. If we view a weak result for some reaction as a separate result from a strong result for the same reaction, then we can say that the phenomenon of additive reaction strengths falls into the class of monotonic abduction problems. That is, each of two antibodies alone might explain a weak reaction. Together, they would explain either a weak reaction or a strong reaction.

Incompatibility. In this domain, some antibodies are incompatible with others. Also, for each antibody, RED distinguishes between two different, incompatible ways that it can react. Thus, red blood antibody identification is clearly outside of monotonic abduction problems and within the intractable class of incompatibility abduction problems. Below, we discuss why this is not usually a difficulty in this domain.

Cancellation. No cancellation interactions take place in this domain.

Ordered. RED rates the plausibility of the presence of an antibody on a 7-point qualitative scale. Because there are about 60 antibody subtypes, the same plausibility rating is given to several antibodies. Strictly speaking, this takes the problem out of ordered abduction problems, but we describe below why this is not usually a problem.

Incompatibility relationships and lack of plausibility ordering do not usually create difficulties in this domain for the following reasons. One is that most antibodies are usually ruled out before any composite hypotheses are considered, i.e., the evidence indicates that the antibodies cannot reasonably be part of any composite hypotheses. The more antibodies that are ruled out, the more likely that the remaining antibodies contain no or few incompatibilities, and resemble an ordered abduction problem.

Another reason is that the reaction testing is designed to discriminate between the antibodies. Thus, an antibody that is present usually explains some reaction more plausibly than any other antibody. An antibody that is not present is unlikely to have clear evidence in its favor and is usually superfluous in the context of equally or higher rated antibodies.

A final reason is that it is rare to have more than a few antibodies. Other antibodies that are rated lower than these antibodies are easily eliminated.

In rare cases, though, these reasons do not apply with the result that RED has difficulties with incompatible pairs or unordered hypotheses, or that RED selects an explanation with many antibodies whereas the preferred answer contains a smaller number of individually less plausible antibodies [27].

7 Discussion

We have discovered one restricted class of abduction problems in which it is tractable to find the best explanation. In this class, there can be no incompatibility relationships or cancellation interactions, the plausibilities of the individual hypotheses are all different from each other, and there must be exactly one best explanation according to the best-small plausibility criterion. Unfortunately, it is intractable to determine whether there is more than one best explanation in ordered abduction problems. However, it is still tractable to find one of the best-small best explanations in ordered monotonic abduction problems.

For abduction in general, however, our results are not encouraging. We believe that few domains satisfy the independent or monotonic property, i.e., they usually have incompatibility relationships and cancellation interactions. Requiring the most plausible explanation appears to guarantee intractability for abduction. It is important to note that these difficulties result from the nature of abduction problems, and not the representations or algorithms being used to solve the problem. *These problems are hard no matter what representation or algorithm is used.*

Fortunately, there are several mitigating factors that might hold for specific domains. One factor is that incompatibility relationships and cancellation interactions might be sufficiently sparse so that it is not expensive to search for explanations. However, only $O(n)$ incompatibilities or cancellations are sufficient to lead to intractability, and the maximum plausibility requirement still remains a difficulty.

Another factor, as discussed in Section 1, is that some constraint might guarantee a polynomial search space, e.g., a limit on the size of hypotheses or sufficient knowledge to rule

out most individual hypotheses. For example, if rule-out knowledge can reduce the number of individual hypotheses from h to $\log h$, then the problem is tractable. It is important to note that such factors do not simply call for “more knowledge,” but knowledge of the right type, e.g., *rule-out* knowledge. Additional knowledge *per se* does not reduce complexity. For example, more knowledge about incompatibilities or cancellations makes abduction harder.

The abductive reasoning of the RED expert system works because of these factors. The size of the right answer is usually small, and rule-out knowledge is able to eliminate many hypotheses. RED is able to avoid exhaustive search because the non-ruled-out hypotheses are close to an ordered monotonic abduction problem.

If there are no tractable algorithms for a class of abduction problems, then there is no choice but do abduction heuristically (unless one is willing to wait for a very long time). This poses a challenge to researchers who attempt to deal with abductive inference—provide a characterization that respects the classic criteria of good explanations (parsimony, coverage, consistency, and plausibility), but avoids the computational pitfalls that beset solutions attempting to optimize these criteria. We believe this will lead to the adoption of a more naturalistic or satisficing conceptualization of abduction [14, 29], in which the final explanation is not guaranteed to be optimal, e.g., it might not explain some data. Perhaps one mark of intelligence is being able to act despite the lack of optimal solutions.

Our results show that abduction, characterized as finding the most plausible composite hypothesis that explains all the data, is generally an intractable problem. Thus, it is futile to hope for a tractable algorithm that produces optimal answers for all kinds of abduction problems. To be solved efficiently, an abduction problem must have certain features that make it tractable, and a reasoning method that takes advantage of those features. Understanding abduction, as for any portion of intelligence, requires a theory of reasoning that takes care for the practicality of computations.

A Proofs of Theorems

In this appendix, we provide a proof for each theorem in the paper. We assume that the functions e and pl are tractable with time complexities $O(C_e)$, and $O(C_{pl})$ in the size of an abduction problem, respectively (see Section 3). The reader is forewarned that many of the reduction proofs do not provide direct insight on the intuitive reasons underlying the complexity results. The proof of Theorem 4.3 is given after Theorem 4.5.

Theorem 4.1 *For the class of independent abduction problems, it is #P-complete to determine the number of explanations.*

This is in #P because each composite hypothesis $H \subseteq H_{all}$ can be generated in nondeterministic polynomial time, and it is easy to check if H is complete and parsimonious.

It is possible to reduce from determining the number of minimal vertex covers [32] to determining the number of explanations. Given a graph G with vertices V and edges E , a minimal vertex cover is a minimal subset of vertices $V' \subseteq V$ such that every edge is connected

to a vertex in V' . An independent abduction problem can be constructed from G as follows:

$$\begin{aligned} D_{all} &= E \\ H_{all} &= V \\ e(H) &= \{(u, v) \in E \mid u \in H \vee v \in H\} \\ pl(H) &= \text{anything} \end{aligned}$$

In this construction, H is an explanation iff it corresponds to a minimal vertex cover.

Theorem 4.2 *For the class of independent abduction problems, there is an $O(n\mathcal{C}_e + n^2)$ algorithm for finding an explanation, if one exists.*

Theorem 4.4 *For the class of monotonic abduction problems, there is an $O(n\mathcal{C}_e + n^2)$ algorithm for finding an explanation, if one exists.*

For these theorems, $n = |D_{all}| + |H_{all}|$. Because the monotonic constraint is more general than the independent constraint, our discussion is oriented for Theorem 4.4. First we consider the correctness of Algorithm 1, and then consider its complexity.

The first conditional determines whether an explanation exists. Because of the monotonic constraint, H_{all} must explain as much or more than any other composite hypothesis. Hence, if H_{all} is not a complete composite hypothesis, then no composite hypothesis is complete. If H_{all} passes this test, then initializing W to H_{all} ensures that W starts off as a complete composite hypothesis.

Within the loop, W remains a complete composite hypothesis because no element is removed if it leads to explaining less than D_{all} . The fact that the result W is also parsimonious can be shown by contradiction. Suppose that $H \subset W$ and $e(H) = D_{all}$. Then, because the problem is monotonic, $H \subseteq H' \rightarrow e(H) = e(H') = e(W)$. In particular, for each $h \in W \setminus H$, $e(H) = e(W \setminus \{h\}) = e(W)$. However, the loop would have removed h from W (or any superset of W) in just this case, implying that $h \notin W$, which is a contradiction. Thus, the loop produces a complete and parsimonious composite hypothesis, i.e., an explanation.

Now consider the complexity of Algorithm 1. We assume sets are represented as bit vectors. Let $k = |D_{all}|$ and $l = |H_{all}|$.

The conditional makes one call to e ($O(\mathcal{C}_e)$) and checks whether its answer is equal to D_{all} ($O(k)$ steps). Assigning H_{all} to W takes $O(l)$ steps.

Since $|H_{all}| = l$, the loop iterates l times. The loop performs l evaluations of e ($O(\mathcal{C}_e)$ each), l set comparisons ($O(k)$ each), and up to l set differences of single elements ($O(1)$ each). Thus the complexity of algorithm is $O(l\mathcal{C}_e + kl + k + 2l)$. Since both k and l are less than n , $O(n\mathcal{C}_e + n^2)$ clearly bounds the complexity of Algorithm 1.

Theorem 4.5 *For the class of independent incompatibility abduction problems, it is NP-complete to determine whether there is an explanation.*

We prove the NP-completeness of this problem by reducing from 3SAT [10]: given a statement in propositional calculus in conjunctive normal form, in which each term has at most three factors, find an assignment of variables that makes the statement true.

Let S be a statement in propositional calculus in 3SAT form. Let $U = \{u_1, u_2, \dots, u_m\}$ be the variables used in S . Let n be the number of terms in S . An equivalent independent incompatibility abduction problem $\mathcal{P} = \langle D_{all}, H_{all}, e, pl, \mathcal{I} \rangle$ can be constructed by:

$$\begin{aligned}
D_{all} &= \{d_1, d_2, \dots, d_n\} \\
H_{all} &= \{h_1, h'_1, h_2, h'_2, \dots, h_m, h'_m\} \\
e(h_i) &= \{d_j \mid u_i \text{ is a factor in the } j\text{th term}\} \\
e(h'_i) &= \{d_j \mid \neg u_i \text{ is a factor in the } j\text{th term}\} \\
pl(H) &= \text{anything} \\
\mathcal{I} &= \{\{h_1, h'_1\}, \{h_2, h'_2\}, \dots, \{h_m, h'_m\}\}
\end{aligned}$$

If H is a complete composite hypothesis for \mathcal{P} , then S is satisfied by the following assignment.

$$u_i = \begin{cases} \text{true} & \text{if } h_i \in H \\ \text{false} & \text{otherwise} \end{cases}$$

Consider the j th term in S . Since $d_j \in e(H)$, there is some i such that either h_i or h'_i is in H and explains d_j . If $h_i \in H$ and $d_j \in e(h_i)$, then u_i is a factor in the j th term of S and by the assignment above, the term is satisfied. If $h'_i \in H$ and $d_j \in e(h'_i)$, then $h_i \notin H$ and $\neg u_i$ is a factor in the j th term of S , so by the assignment above, the term is satisfied. Since j was arbitrary, it must be the case that all terms in S will be satisfied by the above assignment.

If S is satisfied by a value assignment $U' \subseteq U$ indicating:

$$u_i = \begin{cases} \text{true} & \text{if } u \in U' \\ \text{false} & \text{otherwise} \end{cases}$$

then $H = \{h_i \mid u_i \in U'\} \cup \{h'_i \mid u_i \notin U'\}$ will be a complete composite hypothesis for \mathcal{P} .

An explanation for \mathcal{P} exists if and only if a complete composite hypothesis for \mathcal{P} exists. Thus, 3SAT problems reduce to independent incompatibility abduction problems. Incompatibility abduction problems are clearly in NP since it is easy to guess a composite hypothesis H and to test whether $e(H) = D_{all}$. Thus, it is NP-complete to determine whether an independent incompatibility abduction problem has an explanation.

Below, we reduce this class of problems to a number of other classes. For convenience, we shall assume that these abduction problems have the same form as the \mathcal{P} constructed above: the problem is independent except for incompatibilities, each $h \in H_{all}$ is an element of exactly one $I \in \mathcal{I}$, and $|e^{-1}(d)| \leq 3$ for each $d \in D_{all}$. We refer to this special class of independent incompatibility abduction problems as *3SAT abduction problems*.

Theorem 4.3 *For the class of monotonic abduction problems, given a set of explanations, it is NP-complete to determine whether an additional explanation exists.*

This can be reduced from 3SAT abduction problems. Let $\mathcal{P} = \langle D_{all}, H_{all}, e, pl, \mathcal{I} \rangle$ be a 3SAT abduction problem, and let $\mathcal{P}' = \langle D'_{all}, H'_{all}, e', pl' \rangle$ be a monotonic abduction problem constructed from \mathcal{P} as follows.

$$\begin{aligned}
D'_{all} &= D_{all} \\
H'_{all} &= H_{all} \\
e'(H) &= \begin{cases} D_{all} & \text{if } \exists I \in \mathcal{I} (I \subseteq H) \\ e(H) & \text{otherwise} \end{cases} \\
pl'(H) &= pl(H)
\end{aligned}$$

It turns out that \mathcal{I} is a set of explanations for \mathcal{P}' . Consequently, any other explanation can only have at most one hypothesis from each pair $I \in \mathcal{I}$. Thus, it would also be an explanation for \mathcal{P} .

Theorem 4.7 *For the class of diagnosis problems, relative to the complexity of determining whether a composite hypothesis is consistent with $\text{SD} \cup \text{OBS}$, it is NP-complete to determine whether a diagnosis exists.*

To determine whether a solution exists for a diagnosis problem, it is sufficient to exhibit a subset of components $\Delta \subseteq \text{COMPONENTS}$ such that

$$\text{SD} \cup \text{OBS} \cup \{\text{AB}(c) \mid c \in \Delta\} \cup \{\neg \text{AB}(c) \mid c \in \text{COMPONENTS} \setminus \Delta\}$$

is consistent. Hence, diagnosis problems are in NP relative to the complexity of determining whether a composite hypothesis is consistent with $\text{SD} \cup \text{OBS}$.

Determining whether a diagnosis exists can be reduced from finding explanations for 3SAT abduction problems. Let $\mathcal{P} = \langle D_{all}, H_{all}, e, pl, \mathcal{I} \rangle$ be a 3SAT abduction problem, and let $\mathcal{P}' = \langle \text{SD}, \text{OBS}, \text{COMPONENTS} \rangle$ be a diagnosis problem constructed from \mathcal{P} as follows.

$$\begin{aligned} \text{SD} &= \{\text{MANIFEST}(d) \rightarrow \bigvee_{h \in e^{-1}(d)} \text{AB}(h) \mid d \in D_{all}\} \cup \\ &\quad \{\text{AB}(h) \rightarrow \neg \text{AB}(h') \mid \{h, h'\} \in \mathcal{I}\} \\ \text{OBS} &= \{\text{MANIFEST}(d) \mid d \in D_{all}\} \\ \text{COMPONENTS} &= H_{all} \end{aligned}$$

An explanation for \mathcal{P} exists iff there is a complete composite hypothesis H in \mathcal{P} , which is equivalent to whether:

$$\text{SD} \cup \text{OBS} \cup \{\text{AB}(h) \mid h \in H\} \cup \{\neg \text{AB}(h) \mid h \in H_{all} \setminus H\}$$

is consistent for \mathcal{P}' . Clearly, some $\text{AB}(h)$ must be true for each observation $\text{MANIFEST}(d)$. Also, $\text{AB}(h)$ and $\text{AB}(h')$ cannot be true at the same time if $\{h, h'\} \in \mathcal{I}$.

Theorem 4.8 *For the class of belief revision problems, it is NP-complete to determine whether there is a value assignment w to the variables \mathbf{W} such that $P(w|v) > 0$.*

Cooper [4] showed that it is NP-complete to determine whether $P(X = \text{true}) > 0$ for a given variable X . To prove the above theorem, simply let $V = \{X\}$ and v consist of $X = \text{true}$. Determining whether there is a value assignment w such that $P(w|v) > 0$ is equivalent to determining whether $P(X = \text{true}) > 0$.

Theorem 4.9 *For the class of cancellation abduction problems, it is NP-complete to determine whether an explanation exists.*

This can be shown by reduction from 3SAT abduction problems. Let $\mathcal{P} = \langle D_{all}, H_{all}, e, pl, \mathcal{I} \rangle$ be a 3SAT abduction problem. An equivalent cancellation abduction problem

$\mathcal{P}' = \langle D'_{all}, H'_{all}, e', pl', e'_+, e'_- \rangle$ can be constructed by:

$$\begin{aligned}
D'_{all} &= D_{all} \cup \{d_I \mid I \in \mathcal{I}\}^8 \\
H'_{all} &= H_{all} \cup \{h', h''\} \\
e'_+(h) &= \begin{cases} e(h) & \text{if } h \in H_{all} \\ \{d_I \mid I \in \mathcal{I}\} & \text{if } h \in \{h', h''\} \end{cases} \\
e'_-(h) &= \begin{cases} \{d_I \mid h \in I\} & \text{if } h \in H_{all} \\ \emptyset & \text{if } h \in \{h', h''\} \end{cases} \\
pl'(H) &= pl(H)
\end{aligned}$$

Cancellation interactions are created so that each incompatible pair in \mathcal{P} effectively become incompatible in \mathcal{P}' . For all $I \in \mathcal{I}$ and $H' \subseteq H'_{all}$, if $I \subseteq H'$, then $d_I \notin e'(H')$. Thus, no such H' can be an explanation. Hence, \mathcal{P}' has a complete composite hypothesis iff \mathcal{P} has a complete composite hypothesis. In particular, $H \subseteq H_{all}$ is a complete composite hypothesis for \mathcal{P} if and only if $H \cup \{h', h''\}$ is a complete composite hypothesis for \mathcal{P}' . Consequently, \mathcal{P} has an explanation if and only if \mathcal{P}' has an explanation.

Theorem 4.11 *For the class of cancellation abduction problems, it is coNP-complete to determine whether a complete composite hypothesis is parsimonious.*

That is, it is NP-complete to determine whether a complete composite hypothesis is not parsimonious. This can be shown by reduction from 3SAT abduction problems. Let $\mathcal{P} = \langle D_{all}, H_{all}, e, pl, \mathcal{I} \rangle$ be a 3SAT abduction problem, and let $\mathcal{P}' = \langle D'_{all}, H'_{all}, e', pl', e'_+, e'_- \rangle$ be a cancellation abduction problem constructed from \mathcal{P} as follows:

$$\begin{aligned}
H'_{all} &= H_{all} \cup \{h', h'', h^*, h^{**}\} \\
D'_{all} &= D_{all} \cup \{d_I \mid I \in \mathcal{I}\} \cup \{d_h \mid h \in H_{all}\} \\
e'_+(h) &= \begin{cases} e(h) \cup d_h & \text{if } h \in H_{all} \\ \{d_I \mid I \in \mathcal{I}\} & \text{if } h \in \{h', h''\} \\ D_{all} \cup \{d_I \mid I \in \mathcal{I}\} & \text{if } h = h^* \\ \{d_h \mid h \in H_{all}\} & \text{if } h = h^{**} \end{cases} \\
e'_-(h) &= \begin{cases} \{d_I \mid h \in I\} & \text{if } h \in H_{all} \\ \{d_h \mid h \in H_{all}\} & \text{if } h = h^* \\ \emptyset & \text{if } h \in \{h', h'', h^{**}\} \end{cases} \\
pl'(H) &= pl(H)
\end{aligned}$$

This construction is similar to the previous one except that additional data and hypotheses are included so that H'_{all} is a complete composite hypothesis. However, any other complete composite hypothesis (which obviously must be a proper subset of H'_{all}) cannot include h^* and must satisfy cancellation interactions equivalent to \mathcal{P} 's incompatibilities. In particular, $H \subseteq H_{all}$ is a complete composite hypothesis for \mathcal{P} if and only if $H \cup \{h', h'', h^{**}\}$ is a complete composite hypothesis for \mathcal{P}' .

Theorem 5.1 *For the class of independent abduction problems using the best-small plausibility criterion, it is NP-hard to find a best explanation.*

⁸This means that D'_{all} has a datum corresponding to each $I \in \mathcal{I}$, notated as d_I .

This can be shown by reduction from 3SAT abduction problems. Let $\mathcal{P} = \langle D_{all}, H_{all}, e, pl, \mathcal{I} \rangle$ be a 3SAT abduction problem, and let $\mathcal{P}' = \langle D'_{all}, H'_{all}, e', pl' \rangle$ be an independent abduction problem using best-small constructed from \mathcal{P} as follows:

Let f_1 be a function from \mathcal{I} to H_{all} , such that $\forall I \in \mathcal{I} (f_1(I) \in I)$, i.e., f_1 chooses one hypothesis from each pair in \mathcal{I} . Let H_1 be the set of hypotheses that f_1 chooses, i.e., $H_1 = \bigcup_{I \in \mathcal{I}} \{f_1(I)\}$. Let f_2 be another function from \mathcal{I} to H_{all} , such that f_2 chooses the other hypothesis from each pair in I . Now define \mathcal{P}' as:

$$\begin{aligned} D'_{all} &= D_{all} \cup \{d_I \mid I \in \mathcal{I}\} \cup \{d'\} \\ H'_{all} &= H_{all} \cup \{h', h''\} \\ e'(h) &= \begin{cases} e(h) \cup \{d_I \mid h \in I\} & \text{if } h \in H_{all} \\ \{d'\} \cup D_{all} \setminus e(H_1) & \text{if } h = h' \\ \{d'\} & \text{if } h = h'' \end{cases} \\ \forall I \in \mathcal{I} & (pl'(f_1(I)) = pl'(f_2(I))) \\ \forall h \in H_{all} & (pl'(h) < pl'(h') < pl'(h'')) \end{aligned}$$

The remaining orderings of pl' do not matter. Let $n = |\mathcal{I}|$. Note that one hypothesis out of each $I \in \mathcal{I}$ must be chosen to explain all the d_I . Either h' or h'' must be chosen to explain d' . Hence, explanations must be of size $n + 1$ or larger. Now $H = H_1 \cup \{h'\}$ is an explanation of size $n + 1$, where $n = |\mathcal{I}|$. Because h' makes h'' superfluous, any other explanation of greater size must include more than n elements of H_{all} . According to best-small, this would match h' of H against a lower ranking hypothesis, so H must be better than any larger explanation. No other explanation can be smaller, so to get a better explanation according to best-small, h'' must be chosen, and only one hypothesis out of each $I \in \mathcal{I}$ can be accepted so that they explain D_{all} , i.e., a solution to the 3SAT abduction problem \mathcal{P} . Such an explanation would be a best explanation and also show that H is not a best explanation. Hence, finding the best explanation would solve the 3SAT abduction problem. Thus, finding a best explanation for independent abduction problems using the best-small plausibility criterion is NP-hard.

Theorem 5.2 *For the class of belief revision problems restricted to OR interactions, it is NP-hard to find the most probable explanation.*

We reduce from finding a best explanation in independent abduction problems using the best-small plausibility criterion. Let $\mathcal{P} = \langle D_{all}, H_{all}, e, pl \rangle$ be an independent abduction problem where pl satisfies the best-small plausibility criterion. A belief revision problem \mathcal{P}' that preserves the orderings among complete composite hypotheses determined by pl , but

might create additional orderings, is:

$$\begin{aligned}
\mathbf{W} &= \{X_d \mid d \in D_{all}\} \cup \{X_h \mid h \in H_{all}\} \\
\mathbf{V} &= \{X_d \mid d \in D_{all}\} \\
\mathbf{S}_d &= \{X_h \mid d \in e(h)\} \\
\mathbf{S}_h &= \emptyset \\
\mathbf{v} &= \{X = true \mid X \in \mathbf{V}\} \\
\forall s_d ((s_d \rightarrow \exists X_h (d \in e(h) \wedge X_h = true)) \rightarrow P(X_d = true \mid s_d) = 1) \\
\forall s_d ((s_d \rightarrow \forall X_h (d \in e(h) \rightarrow X_h = false)) \rightarrow P(X_d = true \mid s_d) = 0) \\
\forall X_h (0 < P(X_h = true) < .5) \\
\forall X_h, X_{h'} (pl(h) < pl(h') \rightarrow P(X_h = true) < P(X_{h'} = true)) \\
\forall X_h, X_{h'} (pl(h) = pl(h') \rightarrow P(X_h = true) = P(X_{h'} = true))
\end{aligned}$$

Note that $P(X_d = true) = 1$ iff there is some h such that $d \in e(h)$ and $P(X_h = true) = 1$. The network thus consists solely of OR interactions. As a consequence, the conditional probabilities can be concisely represented. This ensures that the size of \mathcal{P}' is polynomial in the size of \mathcal{P} .

For convenience, we will use $P(H|D)$ to denote the probability that X_h is *true* for $h \in H$ and X_h is *false* for $h \in H_{all} \setminus H$, given that X_d is *true* for $d \in D$ and X_h is *false* for $d \in D_{all} \setminus D$. If a value assignment w to all the variables \mathbf{W} assigns *true* to all X_d and to only X_h such that $h \in H$, it is easy to verify that $P(w|\mathbf{v}) = P(H|D_{all})$.

If H is a complete composite hypothesis for the abduction problem \mathcal{P} , then H can be compared to other complete composite hypotheses in the belief revision problem \mathcal{P}' using the expression:

$$\prod_{h \in H} P(h) \prod_{h \in H_{all} \setminus H} (1 - P(h))$$

where $P(h)$ denotes $P(X_h = true)$. By Bayes' theorem:

$$P(H|D_{all}) = \frac{P(H)P(D_{all}|H)}{P(D_{all})}$$

Because of the way the conditional probabilities are set up, H implies D_{all} , so $P(D_{all}|H) = 1$. Because $P(D_{all})$ will always be in the denominator, it is sufficient to compare $P(H)$ with $P(H')$ if H' is another complete composite hypotheses. $P(H)$ is calculated by the expression given above. Obviously $P(H|D_{all}) > 0$ if H is a complete composite hypothesis.

If H is not a complete composite hypothesis, then $P(D_{all}|H) = 0$, and so $P(H|D_{all}) = 0$.⁹

Suppose H^* is the composite hypothesis that corresponds to the MPE w^* for the belief revision problem. To show that H^* is a best explanation, we need to show that H^* is complete, H^* is parsimonious, and that no other explanation H is better than H^* based on the best-small plausibility criterion. From the above discussion, it should be clear that H^* is complete.

Now if H^* were not parsimonious, then some $h \in H^*$ is superfluous, i.e., $H^* \setminus \{h\}$ is complete. However, because $1 - P(h) > P(h)$, it would follow that $P(H^* \setminus \{h\}) > P(H^*)$, which contradicts the fact that H^* corresponds to the MPE. Thus, it must be the case that H^* is parsimonious. Since H^* is also complete, H^* is an explanation.

⁹Or undefined if no complete composite hypothesis exists, in which case, no explanation or MPE exists.

Finally, suppose that another explanation H is better than H^* according to the best-small plausibility criterion. Then there exists a 1-1 function m from H to H^* that satisfies the following conditions: for each $h \in H$, $P(h) \geq P(m(h))$; and either H is smaller than H^* , or there exists an $h \in H$ such that $P(h) > P(m(h))$.

Using the function m , a 1-1 function m' from H_{all} to H_{all} can be constructed as follows:

$$m'(h) = \begin{cases} m(h) & \text{if } h \in H \\ m^{-n}(h) & \text{if } h \in H^* \setminus H \wedge m^{-n}(h) \in H \wedge m^{-(n+1)}(h) \text{ does not exist} \\ h & \text{otherwise} \end{cases}$$

The domain of m is mapped into the range of m . Whatever is in m 's range, but not in m 's domain, is inversely mapped into elements in m 's domain, but not in m 's range. Everything left over is neither in m 's range nor domain, and is mapped to itself.

Because of the best-small constraint, the following can be shown for m' :

$$\begin{aligned} P(h) &\geq P(m'(h)) && \text{if } h \in H \\ 1 - P(h) &> P(m'(h)) && \text{if } h \in H^* \setminus H \wedge \exists h' \in H (m(h') = h) \\ 1 - P(h) &= 1 - P(m'(h)) && \text{otherwise} \end{aligned}$$

Since this mapping matches the factors of $P(H)$ to those of $P(H^*)$, it shows that $P(H) \geq P(H^*)$. Furthermore, best-small guarantees that either $P(h) > P(m(h))$ for some $h \in H$ or that H is smaller, implying there is some $h \notin H$ such that $1 - P(h) > P(m'(h))$. Thus, $P(H) > P(H^*)$.

However, this contradicts the fact that H^* is the MPE. Thus, it must be the case that the MPE for the belief revision problem \mathcal{P}' corresponds to a best explanation for the abduction problem \mathcal{P} . Because finding a best explanation is NP-hard, it is also the case that finding the MPE is NP-hard, even if the belief network is restricted to OR interactions.

Theorem 5.3 *For the class of ordered monotonic abduction problems using the best-small plausibility criterion, there is an $O(n\mathcal{C}_e + n\mathcal{C}_{pl} + n^2)$ algorithm for finding a best explanation.*

where $n = |D_{all}| + |H_{all}|$.

Algorithm 2 is different from Algorithm 1 in only one way: instead of iterating over elements of H_{all} in an arbitrary order in the loop, a specific order is imposed. Thus, by similar arguments as for Theorem 4.4, Algorithm 2 will also find an explanation, if one exists. The change to the loop will result in the addition of at most $O(|H_{all}|)$ evaluations of pl and $O(|H_{all}| \log |H_{all}|)$ steps to sort H_{all} based on pl . Clearly then, the complexity of Algorithm 2 conforms to the order of complexity in the theorem. We now show that the explanation it finds will be a best explanation.

Let W be the explanation it returns. Suppose that, according to the best-small plausibility criterion, there is a better explanation H . Then $|H| \leq |W|$ and there must be a match m of H 's elements to W 's elements, such that H 's elements are just as or more plausible than W 's. The matching hypotheses from H to W cannot all have the same plausibility—that would imply that $H \subset W$ because the abduction problem is ordered. However, because W is parsimonious, no proper subset of W can be complete. Thus, at least one element of H must be more plausible than its match in W .

Suppose that the least plausible element $h_1 \in H$ is more plausible than the least plausible element $w_1 \in W$. This implies that all elements of H are more plausible than w_1 . Because H is an explanation, that would imply that w_1 is superfluous in the context of higher-rated hypotheses, and that the loop in Algorithm 2 would have removed w_1 from the working hypothesis, implying that $w_1 \notin W$, a contradiction. It must be the case then, that $w_1 \in H$, otherwise H would not match up with W .

Suppose that the m least plausible elements of H are exactly the same as those of W . Call this set W_m . Further suppose that H and W differ on their $m + 1$ st least plausible elements. Let h_{m+1} and w_{m+1} be these elements, respectively. Consider the hypotheses $W' \subset H_{all}$ that are more plausible than w_{m+1} . Now by the time that w_{m+1} is considered for removal in the loop, the working hypothesis must be $W_m \cup \{w_{m+1}\} \cup W'$. Since w_{m+1} was not found to be superfluous, it must be the case that $W_m \cup W'$ is not complete. Thus, the $m + 1$ st least plausible element of H cannot be an element of W' . Because $h_{m+1} \neq w_{m+1}$, h_{m+1} must be less plausible than w_{m+1} , which contradicts the supposition that H is a better explanation than W .

Hence, by mathematical induction, no H can be a better explanation than W according to the best-small plausibility criterion. Therefore, for ordered monotonic abduction problems, Algorithm 2 tractably finds a best explanation.

Theorem 5.5 *For the class of ordered independent abduction problems using the best-small plausibility criterion, given a best explanation, it is NP-complete to determine whether an additional best explanation exists.*

This reduction is similar to that of Theorem 5.1. Let $\mathcal{P} = \langle D_{all}, H_{all}, e, pl, \mathcal{I} \rangle$ be a 3SAT abduction problem, and let $\mathcal{P}' = \langle D'_{all}, H'_{all}, e', pl' \rangle$ be an ordered independent abduction problem constructed from \mathcal{P} as follows:

Let f_1 be a function from \mathcal{I} to H_{all} , such that f_1 chooses one hypothesis from each pair in \mathcal{I} . Let H_1 be the set of hypotheses that f_1 chooses. Let f_2 be a function from \mathcal{I} to H_{all} that chooses the other hypothesis from each pair in \mathcal{I} . Now define \mathcal{P}' as:

$$\begin{aligned} D'_{all} &= D_{all} \cup \{d_I \mid I \in \mathcal{I}\} \cup \{d', d''\} \\ H'_{all} &= H_{all} \cup \{h', h'', h^*\} \\ e'(h) &= \begin{cases} e(h) \cup \{d_I \mid h \in I\} & \text{if } h \in H_{all} \\ \{d'\} \cup D_{all} \setminus e(H_1) & \text{if } h = h' \\ \{d''\} & \text{if } h = h'' \\ \{d', d''\} & \text{if } h = h^* \end{cases} \\ \forall I \in \mathcal{I} & (pl'(f_1(I)) > pl'(f_2(I))) \\ \forall h \in H_{all} & (pl'(h) < pl'(h^*) < pl'(h'') < pl'(h')) \end{aligned}$$

The remaining orderings for pl' do not matter. Now $H = H_1 \cup \{h', h''\}$ is a best explanation of size $n + 2$, where $n = |\mathcal{I}|$. Because one hypothesis out of each pair in \mathcal{I} (a total of n hypotheses) is needed to explain $\{d_I \mid I \in \mathcal{I}\}$, and because H includes the more plausible hypothesis of each pair and the two most plausible hypotheses overall, no other explanation of size $n + 2$ or greater can be as good as H . Hence, to construct another best explanation, h' and h'' must be excluded, h^* must be included, and only one hypothesis out of each pair in \mathcal{I} can be accepted, i.e., a solution to the 3SAT abduction problem \mathcal{P} . Thus, another best explanation for \mathcal{P}' exists only if \mathcal{P} has an explanation.

References

- [1] D. Allemang, M. C. Tanner, T. Bylander, and J. R. Josephson. On the computational complexity of hypothesis assembly. In *Proc. Tenth Int. Joint Conf. on Artificial Intelligence*, pages 1112–1117, Milan, 1987.
- [2] T. Bylander, D. Allemang, M. C. Tanner, and J. R. Josephson. Some results concerning the computational complexity of abduction. In *Proc. First Int. Conf. on Principles of Knowledge Representation and Reasoning*, pages 44–54, Toronto, 1989.
- [3] E. Charniak and D. McDermott. *Introduction to Artificial Intelligence*. Addison-Wesley, Reading, MA, 1985.
- [4] G. Cooper. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42(2-3):393–405, 1990.
- [5] R. Davis and W. Hamscher. Model-based reasoning: Troubleshooting. In H. E. Shrobe, editor, *Exploring Artificial Intelligence*, pages 297–346. Morgan Kaufmann, San Mateo, CA, 1988.
- [6] J. de Kleer and B. C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32(1):97–130, 1987.
- [7] J. de Kleer and B. C. Williams. Diagnosis with behavioral modes. In *Proc. Eleventh Joint Int. Conf. on Artificial Intelligence*, pages 1324–1330, Detroit, 1989.
- [8] D. Dvorak and B. Kuipers. Model-based monitoring of dynamic systems. In *Proc. Eleventh Joint Int. Conf. on Artificial Intelligence*, pages 1238–1243, Detroit, 1989.
- [9] L. Eshelman. MOLE: A knowledge-acquisition tool for cover-and-differentiate systems. In S. Marcus, editor, *Automating Knowledge Acquisition for Expert Systems*, pages 37–80. Kluwer, Boston, 1988.
- [10] M. R. Garey and D. S. Johnson. *Computers and Intractability*. W. H. Freeman, New York, 1979.
- [11] A. Goel, P. Sadayappan, and J. R. Josephson. Concurrent synthesis of composite explanatory hypotheses. In *Proc. Seventeenth Int. Conf. on Parallel Processing*, pages 156–160, St. Charles, IL, 1988.
- [12] J. R. Hobbs, M. Stickel, P. Martin, and D. Edwards. Interpretation as abduction. In *Proc. 26th Annual Meeting of the Assoc. for Computational Linguistics*, pages 95–103, Buffalo, NY, 1988.
- [13] J. R. Josephson, B. Chandrasekaran, J. W. Smith, and M. C. Tanner. A mechanism for forming composite explanatory hypotheses. *IEEE Trans. Systems, Man, and Cybernetics*, 17(3):445–454, 1987.
- [14] J. R. Josephson and A. Goel. Practical abduction. Technical report, Lab. for AI Research, CIS Dept., Ohio State Univ., Columbus, OH, 1990.

- [15] H. J. Levesque. A knowledge-level account of abduction. In *Proc. Eleventh Joint Int. Conf. on Artificial Intelligence*, pages 1061–1067, Detroit, 1989.
- [16] R. A. Miller, J. E. Pople, and J. D. Myers. INTERNIST-I, An experimental computer-based diagnostic consultant for general internal medicine. *New England J. of Medicine*, 307:468–476, 1982.
- [17] R. S. Patil, P. Szolovits, and W. B. Schwartz. Modeling knowledge of the patient in acid-base and electrolyte disorders. In P. Szolovits, editor, *Artificial Intelligence in Medicine*, pages 191–226. Westview Press, Boulder, CO, 1982.
- [18] J. Pearl. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29(3):241–288, 1986.
- [19] J. Pearl. Distributed revision of composite beliefs. *Artificial Intelligence*, 33(2):173–215, 1987.
- [20] C. S. Peirce. Abduction and induction. In J. Buchler, editor, *Philosophical Writings of Peirce*, chapter 11, pages 150–156. Dover, New York, 1955.
- [21] Y. Peng and J. A. Reggia. A probabilistic causal model for diagnostic problem solving. *IEEE Trans. on Systems, Man, and Cybernetics*, 17(2):146–162, 1987. Part II appears in SMC-17(3):395–406.
- [22] D. Poole. Explanation and prediction: An architecture for default and abductive reasoning. *Computational Intelligence*, 5(2):97–110, 1989.
- [23] H. E. Pople. On the mechanization of abductive logic. In *Proc. Third Int. Joint Conf. on Artificial Intelligence*, pages 147–152, Stanford, 1973.
- [24] H. E. Pople. The formation of composite hypotheses in diagnostic problem solving: An exercise in synthetic reasoning. In *Proc. Fifth Int. Joint Conf. on Artificial Intelligence*, pages 1030–1037, Cambridge, MA, 1977.
- [25] J. A. Reggia, D. S. Nau, and P. Wang. Diagnostic expert systems based on a set covering model. *Int. J. Man-Machine Studies*, 19(5):437–460, 1983.
- [26] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1):57–95, 1987.
- [27] J. W. Smith, J. R. Josephson, M. C. Tanner, J. R. Svirbely, and P. Strohm. Problem solving in red cell antibody identification: RED’s performance on 20 cases. Technical report, LAIR, CIS Dept., Ohio State Univ., Columbus, OH, 1986.
- [28] J. W. Smith, J. R. Svirbely, C. A. Evans, P. Strohm, J. R. Josephson, and M. C. Tanner. RED: A red-cell antibody identification expert module. *J. of Medical Systems*, 9(3):121–138, 1985.

- [29] J. Sticklen. Distributed abduction. In R. Huber, C. Kulikowski, J. M. Krivine, and J. M. David, editors, *Artificial Intelligence in Scientific Computation: Towards Second Generation Systems*. J. C. Baltzer, Basel, Switzerland, 1989.
- [30] P. Struss and O. Dressler. “Physical negation”—Integrating fault models into the general diagnostic engine. In *Proc. Eleventh Joint Int. Conf. on Artificial Intelligence*, pages 1318–1323, Detroit, 1989.
- [31] P. Thagard. *Computational Philosophy of Science*. MIT Press, Cambridge, MA, 1988.
- [32] L. G. Valiant. The complexity of enumeration and reliability problems. *SIAM J. of Computing*, 8(3):410–421, 1979.