Who chooses the assumptions? *

David Poole†
Department of Computer Science,
University of British Columbia,
Vancouver, B.C., Canada V6T 1Z4
poole@cs.ubc.ca

February 14, 1994

Abstract

In this paper we show how a number of different formulations of nonmonotonic reasoning, probabilistic reasoning and design can be combined into a coherent logic-based abductive framework. This framework is based on allowing consistent assumptions to be used to prove a goal. Different frameworks are characterised by who chooses the assumptions, whether an adversary chooses the assumptions, nature chooses the assumptions, or one gets to choose whatever assumptions one likes.

1 Introduction

In artificial intelligence over the last decade there has been much work in logic-based nonmonotonic, probabilistic and abductive reasoning (see e.g., papers in [17, 44, 26]). In this paper, we show how a particular simple form of abductive reasoning can gives a unifying theme to many seeming disparate reasoning schemes, for example Circumscription [24] and Bayesian

*This paper is to appear in P. O'Rorke (Ed.) *Abductive Reasoning*, MIT Press, 1994.
†Scholar, Canadian Institute for Advanced Research

networks [27]. This can be (and has been) used for such diverse applications as diagnosis, user modelling, recognition and planning [32, 29, 31].
The context in which this is placed is in the area of assumption-based logical reasoning, that has been associated with Pierce's notion of abduction. Pierce's notion was of a rule in inference that said, given g and $a \Rightarrow g$, infer a. This has typically[1] been interpreted in terms of assumption-based reasoning, namely that a is a consistent assumption that can be used to derive g. The term 'abduction' has sometimes been used exclusively for the case where g is an observation, [30], and sometimes for the more general idea that places no restriction on the status of g. This paper is about different specializations of the assumption-based reasoning framework corresponding to different restrictions on the status of a and g.
We consider three different tasks that can be placed into this framework:

Design / Planning
: In the area of design or planning [13], g is a design goal to be achieved, and a is a set of building blocks of the design. Thus we hypothesise a design that provably fulfills its goal. We can use any criteria to choose a design; we can choose one design over another because we happen to like it better.

Recognition / Diagnosis
: A different class of problems arises when g is an observation, and we would like to hypothesise what is in the world that could have produced this observation [30, 32, 40]. It is not up to us to choose the assumptions — 'nature' has already chosen what is true; it is our job to determine which of these explanations is right. We want to determine what is in the world or inside a patient or system that could have produced the observations. We also consider making tests to determine which explanation is correct [43, 7].

Default Reasoning
: A third class of problems is when we do not know whether g is true, but g is something we may want to predict based on assumptions of normality [29, 30, 32]. If we want to be conservative in our predictions, we only predict what we can reach even if an 'adversary' gets to choose the assumptions.

[1] See, for example, the papers in [26].

We first go on to define these notions, and show that they are very closely related to some seemingly disparate areas of recent AI research. These are seen as assumption based reasoning, but differ in who chooses the assumptions: one's self, nature or an adversary. This then is related to the theory of games and economic behaviour [46], in order to see how the current AI theories can be expanded.

2 The abductive framework

The formulation of abduction used is that of Theorist [37, 28], but the formulation has become common (see for example, papers in [26]).
We assume a standard first-order language, using the normal logical connectives such as negation, disjunction, conjunction, implication and quantification [9, 16]. A closed formula is one in which every variable is quantified. An open formula is one where some of the variables are free (not in the scope of any quantifier). A ground formula is one that does not contain any variables. An assumption-based scheme is a pair $\langle F, H \rangle$ where

F is a set of closed formulae called the 'Facts',

H is a set of (possibly open) formulae called the 'assumables' or the 'possible hypotheses'. Let H' be the set of ground instances of elements of H.

Definition 2.1 *A* **scenario** *of $\langle F, H \rangle$ is a subset*[2] *D of H' such that $F \cup D$ is consistent.*

Definition 2.2 *If g is a ground formula, an* **explanation** *of g from $\langle F, H \rangle$ is a scenario of $\langle F, H \rangle$ that together with F implies g.*

Thus, if g is a closed formula, an **explanation** of g from $\langle F, H \rangle$ is a set D of elements of H' such that

- $F \cup D \models g$ and
- $F \cup D \not\models \mathit{false}$.

[2] We treat the set of formulae as the conjunction of the formulae. Whether we mean the set or the conjunction will be clear from the context.

The first condition says that, if D were true so would g, and the second says that D is possible given what is known.

Definition 2.3 *An* **extension** *of $\langle F, H\rangle$ is the consequences of F together with a maximal (with respect to set inclusion) scenario of $\langle F, H\rangle$.*

Thus an extension is made by making as many assumptions (from H) as possible.

Lemma 2.4 *[28] A closed formula is in an extension if and only if it has an explanation.*

Definition 2.5 *A* **minimal explanation** *of g is an explanation of g such that no strict subset is also an explanation of g.*

2.1 Implementation

There are two common ways of implementing explanations: bottom-up or top down.
An ATMS [5] is a bottom up abduction engine where the facts are ground Horn clauses (i.e., consist of definite clauses and integrity constraints). Definite clauses are rules where a conjunction of atoms imply an atom and integrity constraints are rules that imply *false*. The idea of the ATMS is to, for each atom, keep a set (*label*) of the minimal explanations (*environments*) found for that atom. The ATMS forward chains on the rules to find minimal explanations for the atoms at the heads of the rules. Integrity constraints are used to rule out inconsistent sets of assumptions (*nogoods*). Non-minimal explanations are also pruned.
Top-down explanation finding (e.g., [13, 37, 33]), works by backward chaining from what we are trying to explain, collecting the sets of assumptions that were needed in the proof. These are shown to be consistent, by failing to prove they are inconsistent.

3 Prediction versus Explanation

There are two different dimensions in which the use of the assumption-based framework can be varied. The first is the status of g, whether it is known or

whether it is something to be determined. The second is who gets to choose the explanations; whether we should be able to choose whichever assumptions we like or whether an adversary gets to choose the assumptions or we average over the explanations using, for example, probabilities.

The first dimension is whether g is known or whether it is something to be determined. This difference has been seen most in the difference between abductive diagnosis (where g is the observation, and the explanations form different diagnoses for g) and consistency-based diagnoses (where the observations are part of the facts, the defaults are normality assumptions, and diagnoses correspond to extensions).

Abductive diagnosis is first described by Pople [38]. What I call consistency-based diagnosis was first described in these terms by Reiter [42] and de Kleer and Williams [7]. Reiter showed how the generalized set covering model of abduction [40] can be represented within his framework. Poole [29] shows the equivalence between the completion of a logical notion of abduction and consistency-based diagnosis. This was for simple acyclic theories of a standard form described there. The completion result was extended to hierarchical logic programs by Console et. al. [3]. Konolige [21] developed an equivalence between consistency-based diagnosis and the closure of abductive diagnosis, which works for cyclic propositional theories. The price he pays for this is that there is no local closure of each symptom in terms of its local causes — rather then the closure being modular and local to the rules that imply an effect, his closure is global and takes the whole theory into account. Poole [36] expands on the equivalence for acyclic theories allowing a local closure and arbitrary (limited only by acyclicity) constraints on interactions. All of these results are restricted to causal (or fault) theories.

The terms abductive diagnosis and consistency-based diagnoses were first used in [31]. That paper showed how both of these frameworks can be used for fault models and normality models and for the continuum of cases in between. That paper presented examples that are much more sophisticated than the simple causal theories for which there are adequate formal theories. It was shown that even the logical formulation of a single observation needs to be different for each diagnosis model!

In summary, when there is a causal model of the system, and everything is propositional, then abductive diagnosis and consistency-based diagnosis on the closure produce the same result. If the causal model is acyclic, a local completion can serve as the closure [36]. If the causal model is cyclic a more

global closure is needed [21]. When we get beyond these simple cases very little is known about the relationship (see [31]).
There are very good reasons for keeping the distinction between abductive and consistency-based (predictive) diagnoses (apart from that fact that we do not understand the relationships for cases beyond the simple causal propositional theories).
Csinger and Poole [4] show, in cooperative discourse domains, that when we want to do both recognition and design (we want to recognize the goals behind other's utterances as well as design our own utterances), a *shared information constraint* implies that we should do design by abduction and recognition by prediction or design by prediction and recognition by abduction. If we do exclusively abduction or exclusively prediction, then we need to store more information than we need to in order to support both recognition and design.
When adding probabilities to the assumption-based frameworks (see Section 4.4), if we want to enforce the independence of hypotheses then the logic must be very weak. In particular, the only legal knowledge-base for the consistency-based framework would be the one derived from the abductive framework. We set up the framework so that the completion is valid. It is much easier to understand the causal knowledge and the inference procedures in terms of abduction than in terms of completion. The semantics can, however, be best understood in terms of the completion [35].

4 Who chooses the assumptions?

In this section we consider different activities that can be encompassed by the assumption-based framework:

Design / Planning
: g is a design goal to be achieved.
We can choose the 'best' explanation for our purposes.

Recognition / Diagnosis
: g is an observation about the world.
'Nature' has already chosen which assumptions are true; we can only guess (given our observations) what it is that nature has chosen.

		g	
		given	to be determined
who?	self	abductive design	brave prediction / predictive design
	adversary	sceptical abduction	default prediction
	nature	Probabilistic Horn abduction	

Figure 1: Different frameworks captured by the two dimensions

Default Reasoning

g is something we may want to predict.

We can sceptically predict as though an 'adversary' gets to choose assumptions.

Each of these is considered in turn and is shown to correspond to different reasoning frameworks that have been proposed.
Figure 1 gives a table showing how the two dimensions of the status of g and who chooses the assumptions interact producing different reasoning paradigms.

4.1 Choosing the best assumptions

Abduction has been proposed for planning and design [13]. In such a formulation, the assumables become building blocks of a plan or design, and we explain the design goal. An explanation corresponds to a plan or design. The design provably fulfills the design goal (the explanation implies the goal), and is possible (the explanation is consistent).
If we consider, for example, Green's method for deductive planning [19], and ask what it is that we have proved, when we have generated the plan, it is exactly this: we have proved (based on the domain description) that if we were to execute the steps in the plan that the goal would hold. If we are designing circuits, then we hypothesise components and connections that, are possible, and if put together would fulfill our design goal [13].
When we have a set of possible designs, it is up to us to choose any one of them — we know they all fulfill our goals. We may choose our circuit by which one has the least costly components or which circuit has the least area.

We may choose our plan by how long it will take or by how much effort it requires. Alternatively we may just choose an arbitrary one that is easy to generate.

4.2 Adversary choosing the assumptions

One class of assumption-based reasoning is where an adversary chooses the assumptions [30, 32]. If g is in all extensions, then no matter which assumptions an adversary chooses, we will be able to explain g (either we can prove g or make more assumptions to prove g). If g is not in all extensions, then if an adversary can choose the extension which does not contain g, then we cannot make any consistent assumptions to allow us to prove g. Thus membership in all extensions seems to be the right characterisation of "predict if an adversary chooses the assumptions".
The following theorem is a derivation of a number of results [30, 10, 39, 18, 15].

Theorem 4.1 *The following are equivalent:*

1. *g is in all extensions.*
2. *There is a set $\mathcal{D}$ of explanations of g such that $\neg(\bigvee_{D\in\mathcal{D}} D)$ cannot be explained.*
3. *g is true in all minimal models of F, where the ordering on interpretations is defined by $M_1 <_H M_2$ if the assumptions violated in M_1 are a subset of the assumption instances violated by M_2. That is, $M_1 <_H M_2$ if $\{h \in H' : M_1 \models \neg h\} \subset \{h \in H' : M_2 \models \neg h\}$.*

In this theorem, 1 is what we claimed was the appropriate characterisation of prediction even when an adversary chooses the assumptions. Point 2, is in terms of explanations. This is important as it is explanations that we know how to compute. The best way to see point 2 is in terms of arguments. The set $\mathcal{D}$ is a set of arguments for g for which there is no counter argument which simultaneously argues against each element of $\mathcal{D}$ (see [30]). Point 3 is a semantic notion in terms of minimal models [45], that is related to the circumscriptive [23, 22] notion of minimal models (see section 4.2.1 below).

Proof: $1 \Rightarrow 2$. Let $\mathcal{A}$ be the set of all explanations of g. If C is an explanation of $\neg \bigvee_{C_i \in \mathcal{A}} C_i$, then C can be extended to an extension E, in which g does not appear (as $F \cup C$ implies the negation of every explanation of g). Thus if g is in all extensions, no such E exists.

$2 \Rightarrow 1$. Suppose 2 is true. Given such a $\mathcal{D}$, every extension contains at least one element of $\mathcal{D}$ (otherwise the extension is an explanation of the negation of the disjunct in 2). g follows from $F \cup D$, for all $D \in \mathcal{D}$ thus g is in every extension.

$3 \Rightarrow 1$. Suppose g is not in extension E. E is consistent and does not entail g, so there is a model M of $E \wedge \neg g$. M is a model of F, as $F \subseteq E$. M is minimal, as if there is some $M' < M$, there is some $d \in D'$ such that $d \notin E$, d is consistent with E (as M' is a model of $E \wedge d$), which is a contradiction to the maximality of the extension E. Thus g is not true in all minimal models.

$1 \Rightarrow 3$. Suppose g is not true in minimal model M. Let E be the set of consequences of $F \cup \{d \in D' : M \models d\}$. E is an extension, as E is consistent (M is a model of E), and if some $d \in D'$, $d \notin E$, then $E \models \neg d$ (otherwise $E \wedge d$ has a model M', in which case $M' < M$, a contradiction to the minimality of M). g is not in extension E (as it is not a consequence of E, as it is false in a model of E). $\Box$

4.2.1 Relationship to circumscription

Circumscription [23, 24] is a formalism for nonmonotonic reasoning that is defined in terms of second order logic for minimising some formula. It can be defined in terms of a circumscriptive formula [24] or in terms of minimal models [22].
Circumscription is usually defined in terms of fixed and varying predicates. In the rest of this discussion we assume that all predicates are varying. Fixed predicates can be simulated by minimising the predicate and its negation [6]. The minimal models definition of theorem 4.1 (based on that of Geffner [15], but without priorities) is subtly but importantly different to the minimal models definition of circumscription [22]. The definition above can be seen as a syntactic minimisation: we are choosing a minimization based on the

(syntactic) hypotheses rather than on the (semantic) denotation of these hypotheses, as in Circumscription. We are minimising over the syntactic forms of the models (the sets we are comparing are sets of ground atomic formulae). In circumscription, the minimization is in the semantic domain (minimising over individuals rather than over ground terms).
A model is a triple $\langle D, \phi, \pi \rangle$ where D, the domain, is a set of individuals, ϕ is a mapping that maps each ground term to an element of D, and π maps each n-ary predicate symbol to a subset of D^n (those tuples for which the relation is true).
The above theorem holds when we syntactically minimise. The restrictions placed on the circumscription in the work of [39, 18] are the unique names assumption (every distinct term denotes a different individual — ϕ is 1-1) and domain closure assumption (every individual in the domain is named by some term — ϕ is onto), giving an isomorphism between the syntactic and semantic minimisation.
Theorem 4.1 does not require the unique names assumption. For example, the violation set $\{ab(a), ab(b)\}$ cannot be reduced by making $a = b$. This syntactic minimisation is also why we can minimise equality; the minimisation occurs before the terms have been assigned to individuals. We can thus affect this assignment. When minimising in the semantic domain, the minimisation occurs after terms have been assigned to individuals; thus the semantic minimisation cannot affect equality [12], and the unique names hypothesis is needed. For example, the violation set $\{ab(a), ab(b)\}$ can be reduced by making $a = b$. Without the unique names assumption, from the facts $\{ab(a), ab(b), p(a)\}$ semantically minimizing ab (assuming $\neg ab$), will conclude $p(b)$. The syntactic minimization does not let us conclude this.
One of the things that circumscription can do which syntactic minimization, as defined here, cannot do is to conclude universal conclusions. For example, by minimising $p(x)$, but knowing $p(a)$, circumscription can conclude

$$\forall x \; x \neq a \Rightarrow \neg p(x).$$

The syntactic minimization cannot conclude the universal formula, but can only conclude $\neg p(t)$ for each ground term t that is different to a.
While circumscription seems like the right tools for mathematical problems like induction, I would argue that the syntactic minimisation is the right tool for most modelling of assumptions about the world (i.e., commonsense

reasoning). It can handle equality properly, and is more modest in its conclusions. While it may be sensible for an agent to assume each person they meet is honest, it is not sensible to assume that every person is honest. It is exactly this unreasonable universal conclusion that circumscription forces on us.

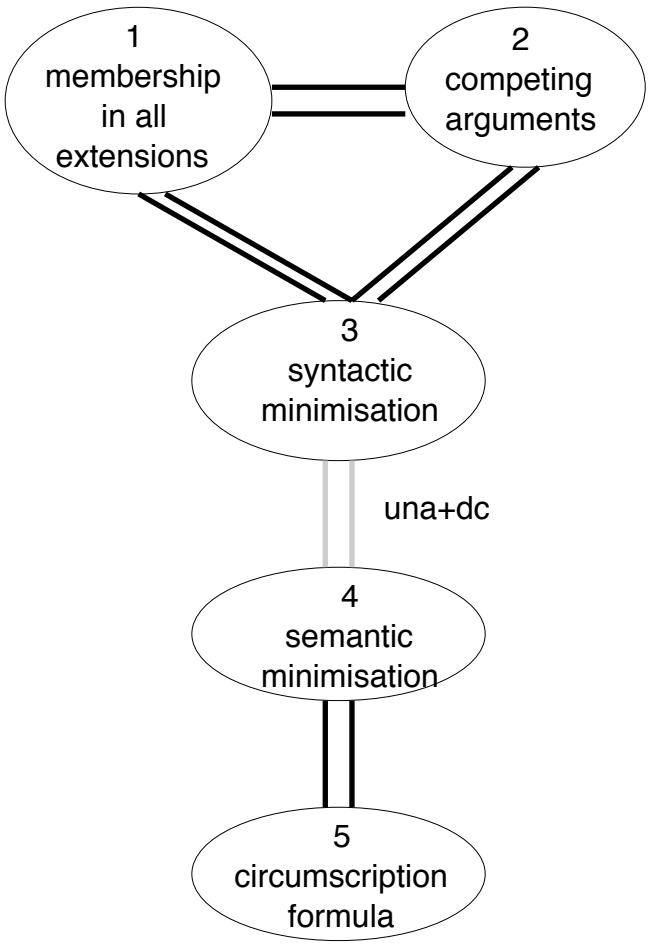
Figure 2 shows the relationship between the formulations of prediction. As well as the above three numbers, 4 denotes the circumscriptive notion of minimisation [23, 22] and 5 denotes the circumscriptive formula. 3 and 4 are the same under the unique names and domain closure assumptions. $1 \Leftrightarrow 5$ is due to Etherington [11, 10]. $1 \Leftrightarrow 2$ is due to Poole [30]. $2 \Leftrightarrow 5$ is due to Przymusinski [39] and Ginsberg [18]. The form of 3 presented here is adapted from Geffner [15], by removing the priorities. $1 \Leftrightarrow 3$, as far as I know, is new to this paper.

4.2.2 Sceptical Prediction Implementations

The idea behind implementing sceptical prediction [30, 20, 39, 18] is that proposition g is in all extensions if it is in an extension even when an adversary can choose the defaults. g is not in all extensions if there is an extension which does not contain g; if we can show that an adversary cannot generate such an extension, then g must be in all extensions.

For the forward chaining default provers [20], to determine if g is in all extensions we try to generate an extension in which g does not appear. When there is a choice of which default to choose, we let an adversary choose the default. If an adversary can generate an extension which does not contain g, then g is not in all extensions. If we can demonstrate that there are no choices for the adversary which lead to an extension not containing g, then g is in all extensions.

For the backward chaining default provers [30, 18, 39], we use the results of Theorem 4.1. Using a method to compute explanations (section 2.1), we find explanations of g and try to find an explanation for the negation of the disjunction of explanations of g. If we fail to find such a counter argument for some set of explanations of g then g is in all extensions. If we find a counter argument to every explanation of g, then g is not in every extension. This can be seen as a form of dialectical argument [30].



deleted by PVO 11/15/93

Figure 2: Relationship between prediction formulations

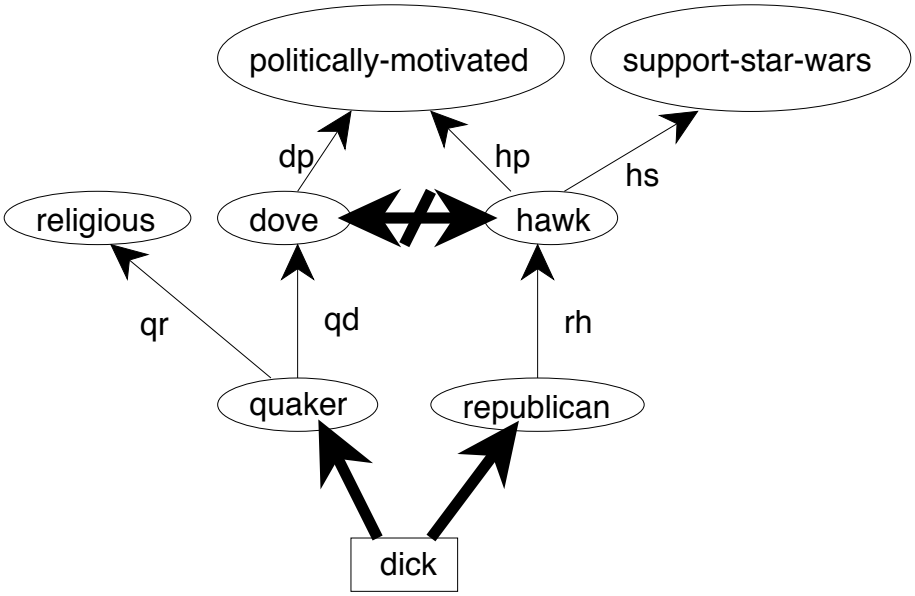


Figure 3: Depiction of Quaker–Republican example.

Example 4.2 *Consider the following example*[3] *depicted in Figure 3:*

$$
\begin{aligned}
H =& \{rh(X), qd(X), hs(X), hp(X), dp(X), qr(X)\} \\
F =& \{\forall X\ republican(X) \wedge rh(X) \Rightarrow \text{hawk}(X), \\
& \forall X\ quaker(X) \wedge qd(X) \Rightarrow \text{dove}(X), \\
& \forall X\ \text{hawk}(X) \wedge hs(X) \Rightarrow \text{support-star-wars}(X), \\
& \forall X\ \text{hawk}(X) \wedge hp(X) \Rightarrow \text{politically-motivated}(X), \\
& \forall X\ \text{dove}(X) \wedge dp(X) \Rightarrow \text{politically-motivated}(X) \\
& \forall X\ \text{quaker}(X) \wedge qr(X) \Rightarrow \text{religious}(X)\} \\
& \forall X\ \neg(\text{dove}(X) \wedge \text{hawk}(X)), \\
& quaker(dick), \\
& republican(dick)\ \}
\end{aligned}
$$

Consider the process of trying to determine support-star-wars($dick$). *There is one explanation for it namely,*

$$F \cup \{rh(dick), hs(dick)\}$$

There is one set of ground instances of defaults which, if an adversary had chosen, would make this argument inapplicable:

$$F \cup \{qd(dick)\}$$

Thus support-star-wars($dick$) *is not in all extensions.*
Consider determining politically-motivated($dick$). *There are two explanations for it:*

$$F \cup \{qd(dick), dp(dick)\}$$

$$F \cup \{rh(dick), hp(dick)\}$$

There is no explanation for the negation of the disjunction of the explanations

$$\neg((qd(dick) \wedge dp(dick)) \vee (rh(dick) \wedge hp(dick)))$$

and so politically-motivated($dick$) *is in all extensions.*

[3]This example is based on an example by Matt Ginsberg, which is based on an example due to Ray Reiter. Here we use the (probably unfortunate) Prolog convention of having variables in upper case.

4.3 Nature choosing the assumptions

The third case is where 'nature' gets to choose the assumptions. In this case, the best we have is a probability distribution over the hypotheses. Probabilistic Horn Abduction [35] is a framework for logic-based abduction that incorporates probabilities with assumptions. This has been implemented [34] and is being used as a framework for diagnosis, user modelling and recognition that incorporates discrete Bayesian Networks [27] as a special case [35].

The aim is to design the knowledge base so that conclusions can be interpreted probabilistically. Associated with each possible hypothesis is a prior probability. Each explanation thus inherits a probability [25], and we build the knowledge base so that the explanations are exclusive and covering. We can then compute the prior probability of any logical expression.

The knowledge base is designed so that the rule base is acyclic and the rules for any goal are disjoint and covering. We use recent results on the completion semantics for abduction [29, 3] that tell us that if the rules for every atom are covering (i.e., Clark's completion [2] holds) then any atom will be equivalent to the disjunction of the explanations for that atom.

We also assume independence amongst consistent hypotheses to allow us to compute the probability of explanations. The idea is that when there is a dependence amongst hypotheses, we invent another hypothesis to explain the dependence. In this manner we can express arbitrary probabilistic dependencies [35]. This idea is essentially Reichenbach's *principle of the common cause* [41].

The probabilistic independence assumption places a restriction on what logic can be used. If we really want different hypotheses to be independent, then we cannot allow the logic to impose any dependence between hypotheses. We cannot allow one hypothesis to entail another or to entail the negation of another. It is for this reason that we restrict the facts to be definite clauses, with a restricted form of integrity constraints.

4.3.1 Probabilistic Horn Abduction

The language is that of pure Prolog (i.e., definite clauses) with special disjoint declarations that specify a set of disjoint hypotheses with associated probabilities. There are some restrictions on the forms of the rules and the

probabilistic dependence allowed.

Definition 4.3 *A* **definite clause** *is of the form: a. or $a \leftarrow a_1 \wedge \cdots \wedge a_n$. where a and each a_i are atomic symbols.*

Definition 4.4 *A* **disjoint declaration** *is of the form*

$$disjoint([h_1 : p_1, \cdots, h_n : p_n]).$$

where the h_i are atoms, and the p_i are real numbers $0 \leq p_i \leq 1$ such that $p_1 + \cdots + p_n = 1$. Any variable appearing in one h_i must appear in all of the h_j (i.e., the h_i share the same variables). The h_i will be referred to as **hypotheses**.

Definition 4.5 *A* **probabilistic Horn abduction theory** *(which will be referred to as a 'theory') is a collection of definite clauses and disjoint declarations such that if a ground atom h is an instance of a hypothesis in one disjoint declaration, then it is not an instance of another hypothesis in any of the disjoint declarations.*

Given theory T, we define the associated facts and hypotheses as:

F_T the **facts**, is the set of definite clauses in T together with the clauses of the form

$$false \leftarrow h_i \wedge h_j$$

where h_i and h_j both appear in the same disjoint declaration in T, and $i \neq j$. Let F'_T be the set of ground instances of elements of F_T.

H_T to be the set of **hypotheses**, the set of h_i such that h_i appears in a disjoint declaration in T. Let H'_T be the set of ground instances of elements of H_T.

P_T is a function $H'_T \mapsto [0, 1]$. $P_T(h'_i) = p_i$ where h'_i is a ground instance of hypothesis h_i, and $h_i : p_i$ is in a disjoint declaration in T.

Where T is understood from context, we omit the subscript.
Probabilistic Horn abduction also contains some assumptions about the rule base. It can be argued that these assumptions are natural, and do not really restrict what can be represented [35].
The first assumption we make is about the relationship between hypotheses and rules:

Assumption 4.6 *There are no rules with head unifying with a member of H.*

Instead of having a rule implying a hypothesis, we invent a new atom, make the hypothesis imply this atom, and all of the rules imply this atom, and use this atom instead of the hypothesis.

Assumption 4.7 *(acyclicity) If F' is the set of ground instances of elements of F, then it is possible to assign a natural number to every ground atom such that for every rule in F' the atoms in the body of the rule are strictly less than the atom in the head.*

This assumption is discussed in [1].

Assumption 4.8 *The rules in F' for a ground non-assumable atom are covering.*

That is, if the rules for a in F' are

$$\begin{array}{c} a \leftarrow B_1 \\ a \leftarrow B_2 \\ \vdots \\ a \leftarrow B_m \end{array}$$

if a is true, one of the B_i is true. Thus Clark's completion [2] is valid for every non-assumable. Often we get around this assumption by adding a rule

$$a \leftarrow some\_other\_reason\_for\_a$$

and making '$some\_other\_reason\_for\_a$' a hypothesis [35].

Lemma 4.9 *[3, 29] Under assumptions 4.6, 4.7 and 4.8, if $expl(g,T)$ is the set of minimal explanations of g from theory T then*

$$g \equiv \bigvee_{e_i \in expl(g,T)} e_i$$

Assumption 4.10 *The bodies of the rules in F' for an atom are mutually exclusive.*

Given the above rules for a, this means that $\neg(B_i \wedge B_j)$ is true in the domain under consideration for each $i \neq j$. We can make this true by adding extra conditions to the rules to make sure they are disjoint.

Lemma 4.11 *Under assumptions 4.6 and 4.10, minimal explanations of atoms or conjunctions of atoms are mutually inconsistent.*

See [35] for more justification of these assumptions.

4.4 Probabilities

Associated with each possible hypothesis is a prior probability. We use this prior probability to compute arbitrary probabilities.
The following is a corollary of lemmata 4.9 and 4.11

Lemma 4.12 *Under assumptions 4.6, 4.7, 4.8 and 4.10, if $expl(g,T)$ is the set of minimal explanations of a conjunction of atoms g from probabilistic Horn abduction theory T then*

$$\begin{aligned} P(g) &= P\left(\bigvee_{e_i \in expl(g,T)} e_i\right) \\ &= \sum_{e_i \in expl(g,T)} P(e_i) \end{aligned}$$

Thus to compute the prior probability of any g we sum the probabilities of the explanations of g.
To compute arbitrary conditional probabilities, we use the definition of conditional probability:

$$P(\alpha|\beta) = \frac{P(\alpha \wedge \beta)}{P(\beta)}$$

Thus to find arbitrary conditional probabilities $P(\alpha|\beta)$, we find $P(\beta)$, which is the sum of the explanations of β, and $P(\alpha \wedge \beta)$ which can be found by explaining α from the explanations of β[4]. Thus arbitrary conditional probabilities can be computed from summing the prior probabilities of explanations. It remains only to compute the prior probability of an explanation D of g. We assume that logical dependencies impose the only statistical dependencies on the hypotheses. In particular we assume:

[4] D is an explanation of $\alpha \wedge \beta$ from $\langle F, H\rangle$ if and only if $D = D_1 \cup D_2$ where D_1 is an explanation of β from $\langle F, H\rangle$ and D_2 is an explanation of α from $\langle F \cup D_1, H\rangle$.

Assumption 4.13 *Ground instances of hypotheses that are not inconsistent (with F_T) are probabilistically independent. That is, different instances of disjoint declarations define independent hypotheses.*

The hypotheses in a minimal explanation are always logically independent. The language has been carefully set up so that the logic does not force any dependencies amongst the hypotheses. If we could prove that some hypotheses implied other hypotheses or their negations, the hypotheses could not be independent. The language is deliberately designed to be too weak to be able to state such logical dependencies between hypotheses.
Under assumption 4.13, if $\{h_1, \cdots, h_n\}$ are part of a minimal explanation, then

$$P(h_1 \wedge \cdots \wedge h_n) \;=\; \prod_{1=1}^{n} P(h_i)$$

To compute the prior of the minimal explanation we multiply the priors of the hypotheses. The posterior probability of the explanation is proportional to this.
Poole [35] shows that all of the numbers can be consistently interpreted as probabilities, and all of the rules can be given their normal logical interpretation.
It can be shown [35] that such a formulation generalises discrete Bayesian networks. The locality of Bayesian networks is preserved in the translation from Bayesian networks to a probabilistic Horn abduction theory.
The mapping is as follows. Suppose random variable a having value v is represented as the proposition $a(v)$. Variable a with parents $b_1, \cdots, b_k$ is translated into the rule:

$$a(V) \leftarrow b_1(V_1) \wedge \cdots \wedge b_k(V_k) \wedge c\_a(V, V_1, \cdots, V_k)$$

where $c\_a(V, V_1, \cdots, V_k)$ is a possible hypothesis. This is a causal hypothesis that says that a has value V because each b_i has value V_i. The probability of this hypothesis is

$$P(a = V | b_1 = V_1 \wedge \cdots \wedge b_k = V_k).$$

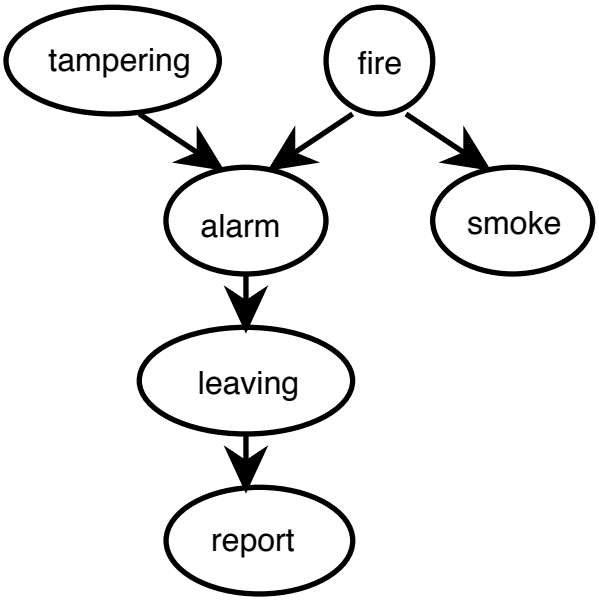
See [35] for details.

Example 4.14 *Consider a representation of the Bayesian network of Figure 4.14, with the following conditional probability distributions:*

$$\begin{aligned}
P(fire) &= 0.01 \\
P(smoke|fire) &= 0.9 \\
P(smoke|\neg fire) &= 0.01 \\
P(tampering) &= 0.02 \\
P(alarm|fire \wedge tampering) &= 0.5 \\
P(alarm|fire \wedge \neg tampering) &= 0.99 \\
P(alarm|\neg fire \wedge tampering) &= 0.85 \\
P(alarm|\neg fire \wedge \neg tampering) &= 0.0001 \\
P(leaving|alarm) &= 0.88 \\
P(leaving|\neg alarm) &= 0.001 \\
P(report|leaving) &= 0.75 \\
P(report|\neg leaving) &= 0.01
\end{aligned}$$

The following is a probabilistic Horn abduction representation of this Bayesian network (from [35]):

$$\begin{aligned}
&disjoint([fire(yes) : 0.01, fire(no) : 0.99]). \\
&smoke(Sm) \leftarrow fire(Fi), c_smoke(Sm, Fi). \\
&disjoint([c_smoke(yes, yes) : 0.9, \\
&\quad c_smoke(no, yes) : 0.1]). \\
&disjoint([c_smoke(yes, no) : 0.01, \\
&\quad c_smoke(no, no) : 0.99]). \\
&disjoint([tampering(yes) : 0.02, \\
&\quad tampering(no) : 0.98]). \\
&alarm(Al) \leftarrow fire(Fi), tampering(Ta), \\
&\quad c_alarm(Al, Fi, Ta). \\
&disjoint([c_alarm(yes, yes, yes) : 0.50, \\
&\quad c_alarm(no, yes, yes) : 0.50]). \\
&disjoint([c_alarm(yes, yes, no) : 0.99,
\end{aligned}$$



deleted by PVO 11/15/93

Figure 4: A Bayesian network for a smoking alarm.

$$
\begin{aligned}
&\quad c\_alarm(no, yes, no) : 0.01]).\\
&disjoint([c\_alarm(yes, no, yes) : 0.85,\\
&\quad c\_alarm(no, no, yes) : 0.15]).\\
&disjoint([c\_alarm(yes, no, no) : 0.0001,\\
&\quad c\_alarm(no, no, no) : 0.9999]).\\
&leaving(Le) \leftarrow alarm(Al), c\_leaving(Le, Al).\\
&disjoint([c\_leaving(yes, yes) : 0.88,\\
&\quad c\_leaving(no, yes) : 0.12]).\\
&disjoint([c\_leaving(yes, no) : 0.001,\\
&\quad c\_leaving(no, no) : 0.999]).\\
&report(Le) \leftarrow leaving(Al), c\_report(Le, Al).\\
&disjoint([c\_report(yes, yes) : 0.75,\\
&\quad c\_report(no, yes) : 0.25]).\\
&disjoint([c\_report(yes, no) : 0.01,\\
&\quad c\_report(no, no) : 0.99]).
\end{aligned}
$$

Here $fire(yes)$ corresponds to there being a fire and $fire(no)$ corresponds to there being no fire. $c\_alarm(yes, yes, no)$ is the causal hypothesis that the alarm is ringing because there is a fire and no tampering. The other variables are treated analogously.

5 Pointers for Future Research

The main problem I am currently interested in is how to mix the above reasoning strategies. The representation language I anticipate having is where some assumptions I choose, some assumptions nature chooses, and some assumptions adversaries (or other agents) choose. We may for example consider a design (that we choose) that will work no matter what other assumptions an adversary makes. We may consider a plan (a design that considers time) that works on the average better than some other plan — thus combining me choosing and nature choosing assumptions. This is very reminiscent of what is called *game theory*.

Game theory [46, 14] has a long history that considers moves by ones self, other agents (including adversaries) and nature, that is in some sense rem-

iniscent of the unified framework provided in this paper. The use of the term 'game' here is much richer than that studied in AI text books for games such as chess. These could be described as deterministic (there are no chance moves by nature), perfect information (each player knows the previous moves of the other players), zero-sum (one player can only win by making the other player lose), two-person games. Each of these assumptions can be generalised [46].
One could claim that this semblance is superficial. Nonmonotonic reasoning is concerned with truth; or determining what is true, based on expectations. Game theory is concerned with moves and decisions. Game theory is inextricably concerned with values and utilities which (currently) play no part in nonmonotonic reasoning.
If this analogy is deeper than this, it is interesting to look at what game/decision theorists have considered that could be incorporated into assumption-based reasoning:

- Moves by nature and agents have been considered. The formalisms described in the preceding section only had one form of assumptions. There were not some assumptions that an adversary, some that nature chooses, and some that the agent itself can choose. Game theory allows for multiple moves by different agents and by nature.

- Utility and values play an integral part in decision and game theory. They are not part of nonmonotonic formalisms, although it has been admitted that values do play a part in what assumptions should be made [45, 8]. Utilities have not been explicit, and maybe they need to be so that they can be reasoned about and not compiled into a set of assumptions.

- What information is available to agents when making a decision is also important (as we do not always have 'perfect information'). This plays an important role in game theory and decision theory. The closest related idea in nonmonotonic reasoning is in the fixed predicates in circumscription [24]. These are assumptions that can be assumed true or assumed false by an adversary [6] (i.e., a is fixed means $a \in H$ and $\neg a \in H$, when used for sceptical prediction). The notion of fixed predicates does not come close to the sophistication needed to reason about information availability.

- Game theory also considers n-person games, for arbitrary n. We do not only need to consider adversaries, but maybe many agents with different values, beliefs and goals. It seems as though nonmonotonic reasoning will need to become intertwined with multi-agent reasoning. With multiple agents we can also consider alliances between agents, and communication between agents.

- Game theory also considers that there is a difference between zero-sum and non zero-sum two person games. In a two person zero sum game we can always treat the other player as an adversary. Many situations are not zero sum, and it may be the case that the agents can gain by cooperation.

Each of these issues is important and points to a wealth of future research.

6 Conclusion

This paper has shown how some recent formulations of reasoning can be placed into a framework of assumption-based reasoning, but differ in who chooses the assumptions. We have a framework that incorporated such seemingly disparate ideas as circumscription and Bayesian networks. This view of recent work sheds light on a whole area of combinations of these reasoning strategies where different assumptions are treated differently.
The abductive framework provides for a unified view of many reasoning strategies that is more general than the minimal model approach advocated by Shoham [45]. As well as being able to capture the notion of minimal models, we can also capture probabilistic reasoning (averaging over models, rather than just choosing models), and design tasks.

Acknowledgements

Thanks to Paul O'Rorke for valuable comments on this paper. This research was supported under NSERC grant OGPOO44121, and under Project B5 of the Institute for Robotics and Intelligent Systems.

References

[1] K. R. Apt and M. Bezem. Acyclic programs. *New Generation Computing*, 9(3-4):335–363, 1991.

[2] K. L. Clark. Negation as failure. In H. Gallaire and J. Minker, editors, *Logic and Databases*, pages 293–322. Plenum Press, New York, 1978.

[3] L. Console, D. Theseider Dupre, and P. Torasso. On the relationship between abduction and deduction. *Journal of Logic and Computation*, 1(5):661–690, 1991.

[4] A. Csinger and D. Poole. Hypothetically speaking: Default reasoning and discourse structure. In *Proc. 13th International Joint Conf. on Artificial Intelligence*, pages 1179–1184, Chamberry, France, August 1993.

[5] J. de Kleer. An assumption-based TMS. *Artificial Intelligence*, 28(2):127–162, March 1986.

[6] J. de Kleer and K. Konolige. Eliminating the fixed predicates from a circumscription. *Artificial Intelligence*, 39(3):391–398, 1989.

[7] J. de Kleer and B. C. Williams. Diagnosing multiple faults. *Artificial Intelligence*, 32(1):97–130, April 1987.

[8] J. Doyle. Constructive belief and rational representation. *Computational Intelligence*, 5(1):1–11, February 1989.

[9] H. B. Enderton. *A Mathematical Introduction to Logic*. Academic Press, Orlando, 1972.

[10] D. W. Etherington. *Reasoning with Incomplete Information*. Research Notes in Artificial Intelligence. Pitman, London, 1987.

[11] D. W. Etherington. Relating default logic and circumscription. In *Proc. 11th International Joint Conf. on Artificial Intelligence*, pages 489–494, Milan, Italy, August 1987.

[12] D. W. Etherington, R. E. Mercer, and R. Reiter. On the adequacy of predicate circumscription for closed-world reasoning. *Computational Intelligence*, 1(1):11–15, 1985.

[13] J. J. Finger and M. R. Genesereth. Residue: A deductive approach to design synthesis. Technical Report STAN-CS-85-1035, Department of Computer Science, Stanford University, Stanford, Cal., 1985.

[14] D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, Cambridge Massachusetts, 1992.

[15] H. Geffner. *Default Reasoning: Causal and Conditional Theories.* PhD thesis, Department of Computer Science, UCLA, Los Angeles, California, November 1989.

[16] M. R. Genesereth and N. J. Nilsson. *Logical Foundations of Artificial Intelligence.* Morgan Kaufmann, Los Altos, Cal., 1987.

[17] M. L. Ginsberg, editor. *Readings in Nonmonotonic Reasoning.* Morgan Kaufmann, Los Altos, Cal., 1987.

[18] M. L. Ginsberg. A circumscriptive theorem prover. *Artificial Intelligence*, 39(2):209–230, June 1989.

[19] C. Green. Application of theorem proving to problem solving. In *Proc. 1st International Joint Conf. on Artificial Intelligence*, pages 219–237, Washington, D.C., May 1969.

[20] H. A. Kautz and B. Selman. Hard problems for simple default logics. In H. Levesque R. Brachman and R. Reiter, editors, *Proc. First International Conf. on Principles of Knowledge Representation and Reasoning*, pages 189–197, Toronto, Canada, May 1989.

[21] K. Konolige. Abduction versus closure in causal theories. *Artificial Intelligence*, 53(2-3):255–272, February 1992.

[22] V. Lifschitz. Computing circumscription. In *Proc. 9th International Joint Conf. on Artificial Intelligence*, pages 121–127, Los Angeles, CA, August 1985.

[23] J. McCarthy. Circumscription—a form of non-monotonic reasoning. *Artificial Intelligence*, 13(1,2):27–39, 1980.

[24] J. McCarthy. Applications of circumscription to formalizing commonsense knowledge. *Artificial Intelligence*, 28(1):89–116, February 1986.

[25] E. M. Neufeld and D. Poole. Towards solving the multiple extension problem: combining defaults and probabilities. In *Proc. Third Workshop on Reasoning with Uncertainty*, pages 305–312, Seattle, July 1987.

[26] P. O'Rorke, editor. *Working Notes, AAAI Spring Symposium on Automates Deduction*. Stanford University, 1990.

[27] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.

[28] D. Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36(1):27–47, 1988.

[29] D. Poole. Representing knowledge for logic-based diagnosis. In *International Conference on Fifth Generation Computing Systems*, pages 1282–1290, Tokyo, Japan, November 1988.

[30] D. Poole. Explanation and prediction: an architecture for default and abductive reasoning. *Computational Intelligence*, 5(2):97–110, 1989.

[31] D. Poole. Normality and faults in logic-based diagnosis. In *Proc. 11th International Joint Conf. on Artificial Intelligence*, pages 1304–1310, Detroit, August 1989.

[32] D. Poole. A methodology for using a default and abductive reasoning system. *International Journal of Intelligent Systems*, 5(5):521–548, December 1990.

[33] D. Poole. Compiling a default reasoning system into Prolog. *New Generation Computing Journal*, 9(1):3–38, 1991.

[34] D. Poole. Logic programming, abduction and probability: A top-down anytime algorithm for computing prior and posterior probabilities. *New Generation Computing*, 11(3–4):377–400, 1993.

[35] D. Poole. Probabilistic Horn abduction and Bayesian networks. *Artificial Intelligence*, 64(1):81–129, 1993.

[36] D. Poole. Representing diagnosis knowledge. *Annals of Mathematics and Artificial Intelligence*, 11(1–4):??–??, 1993.

[37] D. Poole, R. Goebel, and R. Aleliunas. Theorist: A logical reasoning system for defaults and diagnosis. In N. Cercone and G. McCalla, editors, *The Knowledge Frontier: Essays in the Representation of Knowledge*, pages 331–352. Springer-Verlag, New York, NY, 1987.

[38] H. E. Pople, Jr. On the mechanization of abductive logic. In *Proc. 3rd International Joint Conf. on Artificial Intelligence*, pages 147–152, Stanford, August 1973.

[39] T. C. Przymusinski. An algorithm to compute circumscription. *Artificial Intelligence*, 38(1):49–73, February 1989.

[40] J. Reggia, D. Nau, and P. Wang. A formal model of diagnostic inference. *Information Sciences*, pages 227–285, 1985.

[41] H. Reichenbach. *The Direction of Time*. University of California Press, Berkeley and Los Angeles, 1956.

[42] R. Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32(1):57–95, April 1987.

[43] A. Sattar and R. Goebel. Using crucial literals to select better theories. *Computational Intelligence*, 7(1):11–22, February 1991.

[44] G. Shafer and J. Pearl, editors. *Readings in Uncertain Reasoning*. Morgan Kaufmann Publishers, San Mateo, Cal., 1990.

[45] Y. Shoham. Nonmonotonic logics: Meaning and utility. In *Proc. 10th International Joint Conf. on Artificial Intelligence*, pages 388–393, Milan, August 1987.

[46] J. Von Neumann and O. Morgenstern. *Theory of Games and Economic Bahavior*. Princeton University Press, Princeton, third edition, 1953.