# Learning to Reason* The Non-Monotonic Case

Dan Roth*

Division of Applied Sciences
Harvard University
Cambridge MA 02138
danr@das harvard edu

Abstract

We suggest a new approach for the study of the non monotonicity of human commonsense reasoning The two main premises that underlie this work are that commonsense reasoning is an inductive phe nomenon and that missing information in the in lcraclion of the agent with Lhe environment may be as informative for future interactions as observed information This intuition is lormahzed and the problem of reasoning from incomplete information is presented as a problem ol learning attribute func lions over a generalized domain

We consider examples that illustrate various aspects of the non monotonicreasoning phenomena which have been used over the years as bench marks for various formalisms and translate them into Learn ing to Reason problems We demonstrate that these have concise representations over the generalized domain and prove that these representations Lan be learned efficiently

The framework developed suggests an opera- tional approach to studying reasoning that is ne\ ertheless rigorous and amenable to analysis We show that this approach efficiently supports reason- ing with incomplete information and at the same lime matches our expectations of plausible patterns of reasoning in cases where other theories do not

This work continues previous works in the Learn ing to Reason framework and supports the thesis that in order to develop a computational account for commonsense reasoning one should study the phenomena of learning and reasoning together

## 1 Introduction

Any theory aiming at understanding *commensence* reason ing the process that humans use to cope with the mundane but complex aspects of the world in evaluating everyday situa lions should account for lhe flexibility adaptability and speed of commonsense reasoning

The major approach in AI to this problem is within the framework of the knowledge based systems It is assumed that the knowledge is given to the system, stored in some representation language with a well defined meaning and that there is some reasoning mechanism used to determine what can be inferred from the sentences in the knowledge base Earlier formalisms in this framework have abstracted the reasoning lask as a deduction task of determining whether a sentence assumed to capture the situational hand is implied from the knowledge base captunngour theory of theworld This abstraction has been criticised by many (e g [Minsky 1975]) on the ground that it cannot support *non monotonic reasoning*

It is widely acknowledged toda> that a large part of our everyday reasoning involves arriving at conclusions that are nol logically entailed by our theory of the world Many conclusions are derived in the absence of sufficient infor mation to deduce ihem This type of reasoning is naturally nonmonotonic since further evidence may torce us to retract the conclusions In light of this many researchers work ing within the abo\e framework have tned to augment the knowledge base and lo modify the inference mechanisms so as to allow reasoning in lhe presence of incomplete infor mation The idea is lo augmeni the true knowledge (facts and rules) we have about the world with a set of assump lions that capture only typical cases These assumptions are called default assumptions or simply *defaults* Within the knowledge-based sysiems approach defaults are slored in the knowledge base along with the other non-default knowl edge The quest is for a reasoning system that given a query responds in a way lhat agrees with what we know about the world and some subset of the default assumptions and al the same lime supports our intuition about a *plausible umdu sion* The process of reasoning with the knowledge and the defaults is called *default reasoning* and numerous formalisms that attempt al acceptable reasoning behavior have been stud led for it (eg [Al 1980 Touretzky 1986 Reiler 1987 Ethenngton 1988 Goldszmidt and Pearl 1991 Pearl 1988 Gcffner 1990])

Computational considerations however render all the for malisms suggested within the knowledge based systems ap- proach apparently inadequate for commonsense reasoning This is true not only for the lask of deduction but also for many otherforms of reasoning which have been developed [Selman 1990 Roth 1993] Of particular interest in this context are the hardness results on default reasoning tasks [Selman 1990 Papadimitiou 1991] where the increase in complexity (rela live to corresponding deduction tasks) is clearly at odds wilh the intuition lhat reasoning with defaults should somehow re

Research supported by NSF grant CCR 92 00884 and by DARPA AFOSR F4962 92 J 0466

duce the complexity of reasoning Moreover many studies in this framework have shown that capturing what people view as plausible patterns of reasoning is not easy (e g [Tourel-*jVyetal* 1987]) Most formalisms in attempting to capture some aspects of * default reasoning give up on others Multi pie levels of specificity of information irrelevant information and conflicting defaults are among the aspects that the various formalisms have found difficult to reconcile

In [Khardon and Roth, 1994b] a new framework for the study of reasoning is introduced The framework incorporates a role for inductive learning within efficient reasoning and ex hibits the importance of studying the learning and reasoning phenomena together The Learning (in order) to Reason ap proach combines the interfaces to the world used by known learning models with the reasoning task and a performance on tenon suitable for it In this framework the intelligent agent is" given access lo her favorite learning interface and is also given a grace period in which she can interact with this interlace and construct her representation ot the world Her performance is measured in a way that makes explicit the dependence of the reasoning performance on the input from the environment In this framework it is shown that through interaction with the world the agent truly gains additional reasoning power over what is possible in the traditional setting In particular rea soning problems that areprovably intractable in the traditional approach are given efficient Learning lo Reason algorithms

Previous works in the Learning lo Reason framework [ Khardon and Roth, 1994b 1995b] have considered reasoning tasks whose functionality is well defined This paper on the other hand considers tasks in which in many cases there is no agreement on what constitutes a plausible outcome

The disagreement we believe is justified We argue here that commonsense reasoning and in particular reasoning in the presence of incomplete information is an inductive phe nomenon when the notion of consistency is al the heart of the formal reasoning system as in most previous approaches inductive phenomena are difficult lo capture

In this paper we extend the Learning to Reason framework Lo deal explicitly with reasoning in the presence ot incomplete information Inspired by the pac learning approach [Valiant 1934] we present the view that the world is very complicated and there is no hope ot acquiring an exact representation ot it our aim should be lo acquire enough information with which to cope effectively in the world In doing so we extract certain regularities from the world and assume that in similar circumstances we can rely on these

Consider for example concluding from the knowledge that Tweety is a bird thai Tweety can fly This conclusion is use ful and is clearly justified in some situations eg when discussing birds in Boston dunng their migralion season A different conclusion will be suggested though by a veiennar lan working in a birds hospital or by someone raised in an ostnch nature reserve Of course, the possible circumstances in which any presumed correct line of reasoning can be de fealed astound and we are doomed to make mistakes when our expenence does not support the curreni situation

The key to the approach we develop is the view that regular ities occur not only in what we observe (e g if all elephants we have seen had a trunk we might think that all elephants have a trunk) bul also in whal we do not observe (e g if in previous experience of flying birds we were not aware of their color when observing a red bird we would predict that it flies) That is missing information in the interaction of the agent and her environment may be as informative as observed information In this paper we formalize this intuition and use it to develop a theory ihat supports efficient reasoning with incomplele information

Our treatment of incomplele information follows a sugges tion made in [Valiant 1994b] While there in an effort lo formalize the notion *of Rationality* a comprehensive view of the phenomena that compnse cognition is presenled here we presenl a more detailed account of reasoning in the pres ence of incomplete information focusing on presenting it as a problem of Learning to Reason

Unlike previous theones of reasoning in the presence of incomplete information we are not interested in providing a theory of defaults but rather a theory of *inference* We show that the representation developed here provides a richer lan guage for dealing with reasoning problems and consequently many default reasoning scenarios with which previous for malisms have struggled, have concise representations in our framework Moreover these representations con be learned efficiently from interaction wilh the environment to yield of licienl Learning lo Reason algorithms

Later in the paper we discuss the relation of this work lo Lhc default reasoning literature Now wc briefly mention some works thai are related lo Lhc approach presented here In [Khardon and Roth 1995b] a Learning lo Reason approach that can deal with partial information is developed and shown to support efficient deduction The interpretation taken there however is not expressive enough to support non monolonic reasoning; In iKhardon and Roth 1995a] a solution lo some restricted cases of the traditional default reasoning problem is suggested using learnable model based representations The approach presenled in [Sehuurmans and Greiner 1994] is closest lo ours in that they study the problem ol learning default rules The reasoning stage however is nol consid cred and presumably is performed by a traditional reasoner and is thus intractable

After presenting the framework we illustrate in Section 3 how various problems in reasoning with defaults are dealt with in our approach In Section 4 we discuss some of the learning issues this framework raises and some extensions of the work presenled here Wc conclude by discussing lhc results and some theoretical and empirical questions our approach raises

## 2 The Framework

We consider a set \ = $\{r_1 \quad r_n)$ of variables each ol which is associated with a world s attribute and can lake the values I or 0 lo indicate whether ihe associated attribute is true or false in the world An agent intends with the world through a se of *d observed* attributes i — (J,, — 'V, -r, = *v,* , J,,, = i.$_d$) (Wc use *x,* to denote attributes u, lo denote the corresponding values and *v* to denote a vector in *\0,* I *}" ) Many of lhe unobserved attributes mighl not be known[1] to the agent and the assignment lo those and lo known attributes that are unobserved is denoted by the special svmbol * In this wav observations arc vectors in {0 1,*}" bul we write ihem by only specifying the observed variables The world W imposes some distribution *D* over

[1] E g the altnhuie ha;. broken_wing need nol be known

$\{0,1,*\}^n$ that governs the occurrences of the observations $v \in \{0,1,*\}^n$ the agent sees In general we assume nothing about the world $W$ nor about $D$ Presumably there are some functional dependencies in $W$ e g $x_1 = x_2 \wedge x_3$, and those are respected by $D$ in the sense that in any observation $v$ drawn according to $D$ if $v_2 = v_3 = 1$ then $v_1 \neq 0$

We assume that for every known attribute $x_j$ the agent maintains an *attribute function* $f_j \quad \{0,1,*\}^{n-1} \rightarrow \{0,1\}$ that defines the dependence of $x_j$ on the other attributes

An attribute function $f_j$ is represented in a way similar to the way we represent Boolean functions over $\{0,1\}^n$ only that the set of values assigned to each attribute is a (non empty) subset of $\{0,1,*\}$ rather then a (non-empty) subset of $\{0,1\}$ as is usually the case For example a conjunction $f$ that depends on the attributes $x_1, x_2, x_3$ can be written as $f \equiv (x_1 = 1) \wedge (x_2 = 0 \text{ or } *) \wedge (x_3 = * \text{ or } 1)$ A DNF representation for $f$ is written as $f = \bigvee_{j=1}^{m}[(x_{s_1} \in s_{i_1}) \wedge (x_{i_2} \in s_{i_2}) \wedge \quad (x_{i_{k_j}} \in s_{i_{k_j}})]$, where $s_k \subset \{0\ 1\ *\}$ A CNF representation is written in a dual manner Clearly every Boolean function $f$ over $\{0\ 1,*\}^n$ can be represented as a DNF and as a CNF and given $v \in \{0,1\ *\}^n$ it is easy to evaluate $f(v)$

Notice that when using attribute function representations there is no need to make assumptions about the world and in particular to assume it is consistent

We use oracles to model the type of interaction the agent has with the world in the spirit of the formal study of learning [Valiant 1984] and the Learning to Reason framework The oracles differ according to the amount and type of information they supply the agent about the world For the purpose of this exposition we assume that all the interactions of the agent with the world are done via observations $v = (v_{i_1}\ v_i \quad v_{i_d})$

We view the following oracle as the main avenue of inter action with the world the type of interaction which occurs in random situations An *Example Oracle* with respect to the probability distribution $D$ on $\{0\ 1,*\}^n$, denoted $EX(D)$ is an oracle that when accessed returns $v \in \{0,1\ *\}^n$ where $v$ is drawn at random according to $D$ As discussed in [Khardon and Roth 1994a], in situations constrained to satisfy some context condition (e g $Q = \{x_1 = \text{we\_are\_in\_Boston}\}$ or $Q = \{x_1 \wedge x_2 \rightarrow x_3\}$) the occurrences of observations is not governed by $D$ but by the distribution $D_Q$ which is the distribution we see by filtering out all those observations that do not satisfy $Q$ (We follow here the formulation suggested in [Valiant 1994b]) We denote this oracle by $EX(D_Q)$

The following oracle can be thought of as an on-line version of the example oracle and is sometimes more suitable for the learning to reason tasks considered here A *Reasoning Query Oracle* for the attribute function $f_j$, with respect to the distribution $D$ denoted $RQ_D(f_j)$ is an oracle that when accessed performs the following protocol with the agent $A$ (1) The oracle picks $v \in \{0\ 1\ *\}^n$ according to $D$, hides the value of $x_j$ and returns it as a query to $A$ (We denote the query by $rq(v\ j = ?)$) (2) The agent $A$ answers 1 or 0 by evaluating $f_j(v)$ (3) The oracle responds by correct or incorrect A reasoning query oracle for a class $\mathcal{F}$ of attribute functions is denoted by $RQ_D(\mathcal{F})$

We denote by $I$ the *interface* available to the agent in a given situation This can be any collection of oracles that represent a reasonable interaction of the agent with the en vironment and might depend on the arbitrary and unknown distribution $D$ over $\{0\ 1,*\}^n$ or some restriction of it $D_Q$ (We exclude $RQ$ from $I$ for notational convenience ) Other oracles considered include (See [Khardon and Roth 1994b 1995b]) a *Membership Query Oracle* for the attribute function $f_j$ (which on input $v \in \{0,1,*\}^{n-1}$ and $j$ returns $f_j(v)$) an *Equivalence Query Oracle* for $f_j$ (which on input $g \quad \{0\ 1,*\}^{n-1} \rightarrow \{0\ 1\}$ determines wh $f_j \equiv g$) $f_j$ *Causal Example Oracle* and others

The learning scenario most appropriate in our case is an on-line scenario (or continuous learning) [Littlestone 1989 Valiant 1994a] Every example received by the algorithm can be used to update many attribute functions in parallel For example if $v \in \{0\ 1,*\}^n$ is supplied by $EX(D)$ and $V_j = I, v_i =0$ than $v$ can be used as a positive example for the attribute function $j$ and a negative example for $f_i$

The reasoning task we consider is a *prediction* task Given $v \in \{0\ 1,*\}^{n-1}$ in which $ij$ is hidden (i e we do not [illegible] a value for $x_j$) the algorithm is required to predict $f_j(v)$ Thus reasoning with respect to an attribute $i_3$ is reduced to evaluating the attribute function $f_j$ on a total vector over $\{0\ 1\ *\}^{n-1}$ and it depends on learning the correct attribute function We consider a query given to the algorithm as if given by the Reasoning Query Oracle $RQ_D(f_j)$ Thus a reasoning error supplies the algorithm information which in turn can be used to improve its future reasoning behavior In doing so the algorithm may use other oracles from $I$ Notice that the queries depend on the distribution $D$ and thus the algorithm improves Us performance faster' in areas of the distribution in which u is queried more For a class F of attribute functions we say that an algorithm solves the reasoning problem $RQ(F)$ if it can answer prediction queries with respect to all attribute functions $f \in \mathcal{F}$

As performance criteria we will use the criteria accepted in computational learning theory (which we do not define here), namely either the pac critenon [Valiant 1984] or the mistake-bound criterion [Littlestone 1989] Since reasoning is efficient given the attribute functions, we can define An algorithm *A is a Probably Approximately Correct Learning to Reason (PAC L2R) {Mistake Bound Learning to Reason (MB L2R))* algorithm for the reasoning problem $RQ(\mathcal{F})$ if there exists a PAC (Mistake-Bound) learning algorithim for the class $T$ given access to $I$ The algorithm is *noise tolerant* when it can tolerate the standard amount of classification noise[2]

## 3 Default Reasoning

The term default reasoning is used in AI for patterns of inference that permit drawing conclusions suggested but not entailed, by the knowledge available to the system More specifically defaullreasoningisageneral approach within the knowledge-based systems framework, for solving the problem of reasoning in the presence of incomplete information This is usually done by augmenting the "true" knowledge the agent is given about the world with a set of default assumptions that capture what is typically the case When presented widi a query, the inference produced should agree with the true

[2]Classification noise [Angluin and Laird 1988] occurs when there is some probability $\eta$ (the *error rate)* thai the label of an example is flipped (from 0 to 1 or vice versa) Most learning algo nthms known can tolerate classification noise with error rate $\eta < 1/2$ [Reams 1993]

world knowledge and some subset of the default assumptions and at the same lime support our intuition about a *plausible conclusion*

Attempts to represent and reason with defaults have en Lountered many problems (e g [Neufeld, 1989 Poole 1989, Geffner 1990]) In many cases, reasoning with accept able defaults lead to unacceptable conclusions Problems occur whenever defaults interact and can be characterized fre quently as problems of distinguishing good defaults from bad ones But reasons for deciding between good and bad defaults vary and in most cases depend on the situ ation No general method exists according lo which one can rank defaults [Geffner, 1990] The only way to fig ure out why and when certain defaults are preferred lo oth- ers is to understand what the defaults say about the world While probabilistic and statistical approaches [Geffner 1990 Bacchus *et al* 1993] present an important step in this direc lion they still suffer from some of the same problems [Geftner 1994] and are infeasible computationally

The approach developed here does not use defaults Raiher it is a theory of *inference* II reasons from a knowledge rcprc sentation into which the incompleteness is compiled via a learning process As we show later in Section 3 I there is no direct mapping between the way default reasoning problems have been traditionally defined and our framework In order to exhibit the advantages of our approach we translate default reasoning problems into Learning to Reason problems Given a default reasoning problem (I c true world knowledge and a set of defaull assumptions) we suggest a scenario of interac tions wilh the world that reflects the type of observations that could have led to this *view* of the world These observations are used to learn an attribute function representation of the world o v $\{0\ 1\ *\}^n$ e n given T query we argue that this representation yields the sought after response The fol lowing convention is used in presenting the dctaull reasoning examples The traditional representation is given as a set $KB$ of knowledgebase rules and a set $\Delta$ of default rules (As usual penguin(x) —► bird(x) means lhat if x is a penguin then x is a bird ) For each problem we presenlaset of observations about the world The observations are elements in $\{0\ 1\ *\}^n$ hul wc present only a subset of the observed attributes which is of interest lo the current example As usual all the unobserved attributes are assigned *

All the examples discussed below have been studied be- fore in the literature The examples or versions of them represent various aspects of lhe non monolonic reasoning phenomena that have been used over the years as bench- marks for various formalisms We do not know of any traditional formalism that can handle in a satisfying way (efficiently or even qualitatively) all the aspects presented by those examples We note though that our first exam ple is a variant of an example considered in [Valiant 1994a] and that all the examples wc consider here could be con sidered also in the *Rationality* framework and be imple mented in principle on the Neuroidal Model [Valiant 1994b 1994a] A (partial) list of papers that have discussed (d subset of) these examples includes [Bacchus *et al* 1993 Ethenng ion, 1988, Geffner, 1990 Reiler 1980 Reiler and G 1981 Selman 1990 Touretzky *el al* 1987]

Example 1 (Basic Example) *Consider the case in which we know lhat penguins are birds penguins do not fly and we have the default assumption birds fly This \s expressed as the set of facts KB =* {penguin(x) → bird(x), penguin(x) → fly(x)} *and the de- fault statement* $\Delta$ = (bird(x) —• fly(x)) *Given this it is reasonable to assume that in all observations* ne *made so far of the world whenever we saw an observation m which the* penguin *attribute was on (set to* 1) *the* bird *attribute was* 1 *as well and the* fly *attribute was set to* 0 *Moreover we have seen observations in which* bird *Has I and fly w-as* I *In those observations* penguin *was never* 1 *That is a plausible sequence of observations could be*

(bird = 1 penguin = 1, fly — 0)
(bird= l.fly = 1)
(bird = 1 fly = 1 red = 1)
(bird= 1 fly = 1, red = 0)
(bird = 1 penguin — 0, fly = 1, has.beak = 1)
(bird = l,fly = l,has_beak = 1)
(bird = l penguin = 1, fly = 0 has.beak = 1)

*Given these observations the attribute function an agent would Keep for* fly *is* $f_{fly}$ = (bird = l) $\wedge$ (penguin = 0 or *) $\wedge$ (has_beak = l or *) *Consider non a query' re gardings, Tweety* rq((bird = 1) flv — ?) *In this case all we know is that Tweety is a bird (lhat is in this observation the only observed attribute is* bird) *and evaluating* $f_{fly}$ *yields the prediction* fly = 1

Along with seeing many observations similar to the above the agent could have also seen a small number of observations like (bird = I fly = 0)[3] The framework supports this even though a deterministic representation is used for the attribute functions These cases are viewed as *classification noise* where the value supplied by $E\ (D)$ for the function $f_{fly}$ is false Therefore in this model the algorithms used to learn atribute functions should tolerate classification noise Since in Section 4 we show lhat this is indeed ihe case we will not incorporate misclassified observations in the next examples

Example2 (Specificity) *Consider the observations dis cussed in Example 1 and assume a que rv about the penguin Tweety* 'q((bird= 1, penguin = l) fly = ') *In this case evaluating* $f_{fly}$ *yields the prediction* fly = 0 *That is we conclude that Tweety does not fly even though Tweety is a bird and birds (when no other more specific information is known) fly*

Example 3 (Irrelevance-I) *Consider the observations dis cussed above and assume a query about the* red bird *Tweety* 7 q((bird= 1 red = 1) fly=?) *Clearly the observations show that lhe attnbuie red ts irrelevant to the function* $f_{fly}$ *and evaluating it therefore yields the prediction* fly(Tweeiy) — 1

*Of course an agent active in a* green *birds nature reserve might be trained on a different set of observations consist ing of (almost) only* green *birds Consequently \he might believe that greenhood' is a necessary property of flying birds thatis she might have* $f_{FLY}$ = (bird = 1) $\wedge$ (green = 1) *as the attribute function for* fly *There is no contradiction here these are exactly the type of reasoning patterns the sought after theory should possess*

Example 4 (Irrelevance-II) *Consider the observations dis cussed above and a query about the penguin Tweety*

[1]Those observations cannot be the majority of the observations seen since still when all we know about Tweety is that it is a bird we think it flies

rq((bird = 1, penguin = 1), has_beak =?) Here prediction is done b\ evaluating $f_{has\_beak}$ Note thai there is no relation between the attribute functions $f_{has\_beak}$ and $f_{fly}$ These are acquired in parallel and the fact that penguins have special properties with respect to flying does not mean they need to have exceptional properties with respect to having a beak Clearly the observations lead to $F_{has\_beak}$ = (bird = 1) and evaluating it yields has_beak = 1

We note that while the conclusion above is very intuitive it is not supported by many treatment v of default reasoning (e g [Kraus et a) 19901) which encounter difficulties in trying to support both specificity and irrelevance

Example 5 (Multiple Extensions) Consider the set of facts KB = {bal(x) —► mammal(x)} and default statements A = {mammal(x) —■ fly(x),bal(x) -» fly(x),dead(x) — fly(x)} Given that it is reasonable to assume that the observations made of the world had the following properties in observa tions with a bat attribute set to 1 the mammal attribute was 1 as well we have observed bais that fly but also mammals that do not fly in the latter case bat was not 1 also we have not seen dead things fly Therefore a plausible set of observations could be

(mammal =1 bat — 1 fly = 1)
(bat= l,fly= 1)
(mammal = 1, fly = 0)
(mammal =1 bat = 0, fly = 0 red = 1)
(dead = 1)
(mammal =1 bat = 0, dead = 1)
(bat = 1 dead = 1)
(bat = l,dead= l, fly = 0}

Her? the attribute function an agent would keep for fly[4] is $fa_y$ = (bat = 1) A (dead = 0 or +) Con sider now a query regarding Dracula presented as rq((bat = I dead = 1) fly -?) Clearly evaluating $f_{fly}$ on this observation yields the prediction fly(Dracula) — 0 In case all we know is that Dmcula is a bat and we do not know that it is dead (that is dead=*) the query is rq(bal = 1 fly =?) and evaluating $f_{fly}$ Melds the prediction fly(Dracula) — 1

As before there is no contradiction here, these are exactly the type of reasoning patterns the sought after theory should pos sess The traditional treatment runs in this case into problems of conflicting defaults For example one has to decide which of the default rules, bat(x) —>■ fly(i) or dead(r) —► fly(i) to apply in order to predict the value of fly(Dracula)

Example 6 (Preferences) Assume the default statements are given by A = {student(x) —■ employed(x), adull(x) — employed(x), student(x) —* adull(x)) andtheset of facts is empty These defaults were written in this way to reflect a situation in which the agent observes the following properties m observations in which the student attribute was set to 1 the employed attribute was not set to 1 in observations in which the student attribute was set to 1 the adult attribute was not set to 0 in observations in which the adull attribute was set to I the employed attribute was not set to 0 unless some other information is given The following observations could have been seen by the agent

(student = 1 employed = 0)
(student = 1, adult = 1)
(employed = 1 adult = 1)
(student = 0 employed = 1, adult = 1)
(student = 1, employed = 0 adull = 1)

Given these observations the attribute function an agent would keep for employed is $f_{employed}$ = (adult = 1) A (student = 0 or *) On the other hand these observations do not give us enough information to support prediction of the attribute adult in a simple way (see below)

Many othier problems can be handled in a natural wav just as the problems considered above In particular this approach suggests a natural solution to the frame problem which is concerned with how to indicate which aspects of [he world do not change when an action takes place [ML Carthy and Hayes, 1969] While the standard non monotonic reasoning formalisms do not capture the desirable behavior that things stay as they are [Hanks and McDermott, 1986] our representation of incomplete information does so [Roth 1995]

What is most striking about these examples is not only the fact that these examples with which various default reason ing formalisms struggle have a unified representation in our framework but even more so

Observation 1 /n all the cases presented above the attribute function for the attribute of interest can be represented as a conjunction over $(0\ I,*)^n$

It is an empirical question whether there are naturally arising reasoning problems in which the sought after aunbute cannol be represented as a simple function over $\{0, 1, *]^n$ It is ex peeled for example that in situations traditionally presented by a large set of interacting defaults the resulting attribute function might be more complicated However even in this case reasoning reduces to function evaluation and is thus computationally easy In Section 4 we show that we can actu ally learn to reason with function classes which are far more expressive than is needed in the examples discussed above

## 3 1 Relations to Other Formalisms

There is no direct mapping between our treatment of in complete information and traditional formalisms for default reasoning As an example consider the case of preferred interpretations [McCarthy 1980 Selman and Kautz 1990 Papadimitriou 19911 There a theory O and a set A of dc faults are given The theory delines a set of possible models and the default rules define a preference relation (a partial order) on those Once a preferred model is found, inference is done by evaluating queries in this model While this for malism leads to some intriguing mathematical problems, we argue iat one need not solve those in order to reason in a way that agrees with the incomplete default information

Consider Example 6 There no minimal model exists that can capture the intuitive inference with respect to all the at tributes Given the observations the attribute function for em- ployed is $f_{employed}$ = (adult = l)A(sludent = 0 or *) These observations however, do not support a conjunction as an al tribute function for adull but rather the following DNF-hke function $f_{aduit}$ = ((employed = 1) A (student = Oor *)) V ((employed — Oor +) A (student = 1)) Therefore in this case using a single model in {0, 1}" to characterize the sit uation, does not support the ' intuitive conclusion (While

[4] Wc could disjunct H with the function from Example 1 but will assume for clarity thai those are different agents

making the problem harder computationally ) Instead our approach uses the available data to learn the situations in which a specific attribute is on This can always he done and the only question remains is how complex is the representation and whether it can be learned efficiently

## 4 Learning to Reason

Reasoning with respect to an attribute $T_j$ is reduced in this framework to evaluating the attribute function $f_3$ on a loLal vector in $\{0\ 1,*\}^{n-1}$ Assume that our attribute functions are in a class $T$ of Boolean functions over $\{0\ 1\ *\}^n$ If we have efficient learning (to classify) algorithms for $T$ that can tolerate classification noise we can Learn to Reason with F

It turns out that many of the existing learning algorithms for Boolean functions studied in computational learning theory (see a survey in [Blum *et al* 1994]) can be extended to learning algorithms over $\{0\ 1\ *\}^n$ Since in all the examples considered in Section 3 we used the oracle $E\backslash(D)$ only we start by considering learning from examples only

We extend the standard elimination algorithm for learning conjunctions LVahanl 19841 lo work over $\{0\ 1\ *\}^n$ In this case the values assigned to the variables arc non empt) sub sets ol (0 1 *} rather than of {0, I} as is usually the case In the usual elimination algorithm the convention is that when a variable *x,* is allowed lo have any value in (0 1} we omit it from the conjunctive representation We use the same conven Lion here Moreover wc use this convention tor variables that have never been observed In order for variables that have not been observed yet (i e never appeared as 0 or 1) not lo appear in the conjunctive representation the algorithm uses Ihc first positive example to initialize its hypothesis From then on it (1) adds lo the conjunction only newly observed allnbules and (2) uses elimination over the set of known attributes It can be shown that this procedure provides a mistake bound andtherefore a pac algorithm for Boolean conjunctions over

Using the techniques introduced in [Kushilevits and Roch 1995] we can show how to learn kDNF and kCNF formulae over $\{0\ 1,*\}^n$ for any fixed $k$ Moreover these algorithm are shown to tolerate noise and thus can be used to construct L2R algorithms To summarize (see [Roth 19951)

Theorem 1 *Let F be the class of conjunctions disjunctions kCNF and kDNF formulae over* $\{0\ 1\ *\}^n$ *Then there exists an efficient and noise tolerant PAC L2R (MB L2R resp ) algorithm for the rasoning problem RQ(f) that uses the example oracle E\(D)(RQ*$_D$*{fj) resp )*

A richer class of functions can be learned when given access to membership queries in addition lo examples [ Angluin 1988 Blum *et al* 1994 Bshouly 1993] Many of these algo nlhms can be extended lo work over {0 1 *}" In particular using the algorithms studied in iBshouiy 1993] we have

Theorem 2 *There exists an efficient PAC L2R algorithm that uses RQD(*$F_J$*) and MQ(*$f_j$*) for the reasoning problem RQ(F) where*
*(i) F is the class of Decision Trees oxer* {0, 1 * )"
*(n)F is the class of log nCNF n DNF over* {0 1 *}"

We have discussed a knowledge representation that consists of a collection of attribute functions Using our interpretation of incomplete information it can be shown [Roth, 1995] that other representations can support the reasoning behavior demonstrated in this paper Consequently different learning questions may arise the reasoning algorithms might be more complicated and one can also pose more general queries[5] In particular il can be shown that the algorithms used in [Khardon and Roth 1994b] lo learn model based rep resentation can be extended to work over {0 1,*}$^n$ Together with the incomplete information mlcrprelation suggested here this yields the sought after non-monotonic behavior

## 5 Discussion

We have presented a new approach lo the problem of reasoning with incomplete information The main premises of our approach are that (I) It views reasoning as an inductive phenomenon by interaction with the environment the intelligent agent inductively learns a representation of the world and uses it lo respond lo queries The perlormance on the reasoning task is measured in a way that makes explicit the dependence of the reasoning performance on the input from the world (2) Missing information in the interaction of the igent with the environment is taken lo be as informative as observed information

Wc have formulated the problem of reasoning with incomplete Inlormaiion as a problem of learning attribute functions over the domain (0, I, *}" This formulation can tolerate observations that arc inconsistent these are handled as noisy input lo the learning algorithm Moreover multiple levels of specificity of information irrelevant information and con flicting observations are handled in a natural way lo yield conclusions thai malch our inluihon These issues determine the complexity of the attribute function representation But, efficient and noise tolcranl learning algorithms exisl even for function classes over {() 1,*}" that arc far more expressive than was required in the bench marks examples considered

We view the large body of research on defeasible theories of reasoning as an attempt lo characterize the type of defeasible reasoning people do While there is today some understand ing of human like patterns of reasoning we believe that no definition can be given for the type of behavior expected given an abstract representation of partial knowledge as a starting point The Learning lo Reason framework suggests an operational approach to studying reasoning that is never theless rigorous and amenable lo analysis As we have argued here it can be shown to malch our expectations in cases in which the reasoning problem is well dehned

This work suggests several areas in which further theoreti cal study is needed as well as some interesting questions for empirical study Studying other forms of interaction in the learning process extending the framework lo a probabilistic domain and efficient learning in the presence of irrelevant attribules are some of the theoretical questions whose study will help develop and substantiate the claims made here

As mentioned before determining how complex the at tribute functions in naturally arising reasoning problems are and whether those can indeed be represented as sjmple functions over {0, 1, * ] $^n$, is an important empirical question Per

[5] More general quenes are queries with respect to more than a single attribute Notice however that the reasoning tasks considered in most of the default reasoning literature are prediction tasks quenes with respect to a single attribute as we do here

haps the major difference between the knowledge-based sys tern approach to reasoning and the Learning to Reason ap proach is that our approach suggests that in order to make theories of reasoning work m practice we need to train them over a large number of examples Therefore, finding good and large tesl beds on which to validate this theory is one of the most important next steps

## References


[AI 1980] AI Special issue on non monoionic logic Artificial Intelligence 13(1 2) 1980

[Angluin and Laird 1988] D Angluin and P Laird Learning from noisy examples Machine Learning 2(4)343-370 1988

[Angluin 1988] D Angluin Queries and concept learning Ma chine Learning 2(4)319-342 Apnl 1988

[Bacchus et al 1993] F Bacchus A Grove J Y Halpem and D Koller Stanstical foundations for defaull reasoning In Pro ceedings of the International Joint Conference of Artificial Intel ligence pages 563-569 1993

[Blum et al 1994] A Blum R Khardon A Kushilevitz L Pitt, and D Rolh On learning read k satisfy DNF In Proceedings of the Annual ACM Workshop on Computational Learning Theory pages 110-117 1994 (Submitted for publication)

[Bshouly 1993] N H Bshouiy Exact learning via the monotone theory In Proceedings of the IEEE Symp on Foundation of Computer Science pages 302-311 Palo Alto CA 1993

[Elhenngton 1988] D W Ethenngton Reasoning With Incomplete Information Morgan Kaufmann 1988

[Geffner 1990] H Geffner Default Reasoning Casual and Con dtional Theories MIT Press 1990

[Geffner 1994] H Geffner Causal defaull reasoning Principles and algorithms In Proceedings of the National Conferece on Artificial Intelligence pages 245-250 1994

[Galdszmidt and Pearl 1991] M Goldszmidl and J Pearl System Z+ A formalism for reasoning with variable strength defaults In Proceedings ofthe National Conference on Artificial Intelligence pages 399-404 J 991

[Hanks and McDermott 1986] S Hanks and D McDermotL De fault reasoning nonmonotonic logics and the frame problem In Proceedings of the National Conference on Aritfii Ial Intelligence pages 328-333 1986

[Keams 19931 M Keams Efficient noise tolerant learning trom statistical queries In Proceedings of the Twenty Fifth Annual ACM Symposium on Theory of Computing pages 392-401 1993

[khardon and Roth 1994a] R Khardon and D Roth Exploiting relevance through model hased reasoning In AAA/ Fall Svmpo sium on Relevance pages 109-114 1994

[Khardon and Roth 1994b] R Khardon and D Roth Learning to reason In Proceedings of the National Conference on Arufi cial Intelligence pages 682-687 1994 Full version TR 02 94 Aiken Computation Lab Harvard University January 1994

[Khardon and Roth 1995a] R Khardon and D Roth Defaull reasoning with models In Proceedings of the International Joint Conference of Artificial Intelligence August 1995

[Khardon and Roth 1995b] R Khardon and D Roth Learning to reason with a restricted view In Workshop on Computational Learning Theor\ luly 1995

[KntusetaL 19901 S Kraus D Lehmann and M Magidor Pref erenual models and cumulative logics Artificial Intelligence 44 167-207 1990

[Kushilevitz and Roth 1995] E Kushilevitz and D Roth On learn ing visual concepts and DNF formulae Machine Learning 1(1) 11-46 1995 Earlier version appeared in Proceedings of the ACM Workshop on Computational Learning Theory 93

[Littlestone 1989] N Littlestone Mistake bounds and logarithmic linear\hreshoidl e a r n i n g algorithms PhD thesis U C Santa Cruz March 1989

[McCarthy and Hayes 1969] J McCarthy and P Hayes Some philosophical problems from the standpoint of artificial intelli gence In B Meltzer and D Michie editors Machine Intelligence 4 Edinburgh University Press 1969

[McCarthy 1980] J McCarthy Circumscriptum - a form of non monotomc reasoning Artificial Intelligence 13(1 2) 1980

[Minsky 1975] M Minsky A framework for representing knowl edge In P Winston editor The Psychology of Computer Visiom McGraw Hill 1975 Also in R Brachman and H Levesque Readings in Knowledge Representation 1985

[Neufeld 1989] E Neufeld Default and probabilities extension', and coherence In Proceedings of the International Conference on the Principles of Knowledge Representation and Reasoning pages 312-323 1989

[Papadimilnou 1991] C H Papadimitnou On selecting a satisfy ing truth assignment In Proc 32nd Ann IEEE Symp on Foun dations of Computer Sciene pages 163—169 1991

[Pearl 1988] J Pearl Probabilistic Reasoning in Intelligent Sys terns Networks oj Plausible Inference Morgan Kaufman 1988

[Poole 1989] D Poole What the lottery paradox tells us aboul default reasoning In Proceedings of'the InternationalConference an the Principles of Knowledge Representation and Reasoning pages 333-340 1989

[Rener andG 1981] R Reiter and Cnscuolo G On interacting defaults In Proceedings of the International Joint Conference of Artificial Intelligence pages 270-276 1981

[Rcitcr 1980] R Reiter A logic tor default reasoning Artificial Inteligence 13(l 2) 1980

[Reiter 1987] R Reiter Nonmonotonic reasoning In Annual Re views of Computer Science pages 147-188 1987

LRoth 1993] D Rolh On the hardness or approximate reasoning In Proceedmgs of the International Joint Conference of Artificial Intelligence pages 613-6l8 August 1993 To Appear in Artificial Intelligence Journal 1995

IRoth 1995] D Roth Learning to reason the non monotomccase 1995 Full Version In Preparation

ISchuurmans and Greiner 1994] D Schuurmans and R Greiner Learning default concepts In Proceedings of the Tenth Cana dian Conference on Artificial Intelligence {CSCSI 94) 1994

[Selman and Kautz 1990] B Selman and H Kauiz Model preference default theories Artificial Intelligence 45 287-322 1990

[Selman 19901 B Selman Tractable Default Reasoning PhD thesis Department of Computer Science University of Toronto 1990

[Toureizky et al 1987] D Toureizky J Horty and R Thomason A clash of intuitions The current state of nonmonotonic multiple inhentance systems In Proceedings of the International Joint Conference of Artificial Intelligence Morgan Kaufman 1987

[Touretzky 1986] D Touretzky The Mathematics of Inheritance Systems Morgan Kaufman 1986

[Valiant 1984] L G Valiant A theory of the learnable Communi cations of the ACM 27(11) 1134-1142 November 1984

[Valiant 1994a] L G Valiant Circuits of the Mind Oxford Uni versity Press November 1994

iValiam 1994b] L G Valiant Rationality Technical Report TR 32 94 Aiken Computation Lab Harvard University November 1994