Artificial Intelligence and Law (2006) 13: 9–51
DOI 10.1007/s10506-006-9004-2

AGATHA: Using heuristic search to automate the construction of case law theories

ALISON CHORLEY and TREVOR BENCH-CAPON
Department of Computer Science, The University of Liverpool, Liverpool, UK
(E-mail: tbc@csc.liv.ac.uk)

Abstract. In this paper we describe AGATHA, a program designed to automate the process of theory construction in case based domains. Given a seed case and a number of precedent cases, the program uses a set of argument moves to generate a search space for a dialogue between the parties to the dispute. Each move is associated with a set of theory constructors, and thus each point in the space can be associated with a theory intended to explain the seed case and the other cases in the domain. The space is large and so an heuristic search method is needed. This paper describes two methods based on A* and alpha/beta pruning and also a series of experiments designed to explore the appropriateness of different evaluation functions, the most useful precedents to use as seed cases and the quality of the resulting theories.

Key words: case based reasoning, heuristic search, legal knowledge based systems, theory construction

1. Introduction

Bench-Capon and Sartor (2003) it was proposed that reasoning with legal cases should be seen as a process of theory construction, the objective of the reasoning being to construct a theory of the domain which explains as many of the existing cases as possible while giving the desired outcome in the current case. That paper defined a theory as a five tuple comprising; a set of cases to be explained; a set of factors (each associated with an outcome and a social value) with which to describe the cases; a set of defeasible rules linking factors to outcomes; a set of preferences between rules; and a set of preferences between social values which are used to explain preferences between rules. The paper also gave a set of constructors which could be used to build a theory. Importantly (Bench-Capon and Sartor 2003) also linked these theory constructors to argument moves as used in leading Case Based Reasoning Systems such as HYPO (Ashley 1990) and CATO (Aleven 1997), so as to preserve a relation between theory construction and legal argument.

In a series of experiments we have been empirically exploring the ideas of Bench-Capon and Sartor (2003). In the Case Theory Editor (CATE) experiments (Chorley and Bench-Capon 2003a, b, c), (Chorley and Bench-Capon 2004a, b), we developed a software tool which facilitated the manual construction of theories and translated them into executable code. These experiments demonstrated the feasibility of producing explanatory theories using the framework of Bench-Capon and Sartor (2003), suggested some principles of theory construction, and explored the effect of using different methods of comparing sets of factors and values. From this platform we developed, Argument Agent for Theory Automation (AGATHA), which was intended to explore the automation of theory construction.

The underlying idea of AGATHA is that it should construct theories as a side effect of producing an argument in the style of HYPO and CATO like systems. Thus AGATHA is provided with a set of argument moves, each associated with a set of theory constructors. Given a seed case, Plaintiff and Defendant take it in turns to play one of these argument moves, and each of these moves has the effect of modifying the developing theory. Thus, at the conclusion of the debate we have a theory, and the case based dialogue which produced it. This gives the benefits of both rule based and case based approaches to the representation of cases. The theory provides an explicit body of the knowledge for inspection, critique and modification while the process of construction can reflect the practice of legal argument.

2. Argument moves in HYPO and CATO

When thinking about how to argue a new case on the basis of case law, it appears that people think in terms of analogising a past case to the problem or by distinguishing an unfavourable case, rather than in terms of the theory constructors proposed in Bench-Capon and Sartor (2003). Therefore we wish AGATHA to operate by following a series of argument moves as found in case based reasoners. We therefore take the moves of HYPO (Ashley 1990) and CATO (Aleven 1997) as our starting point.

HYPO creates 3-ply arguments using the following four moves:

1. Analogising a problem to a past case with a favourable outcome.
2. Distinguishing a case with unfavourable outcome.
3. Citing a more on point counterexample to a case cited by an opponent.
4. Citing an as-on-point counterexample.

Either party may start the argument by using the first move and analogising the problem case to a past case. The opposing party can then use the remaining three moves to distinguish or counter the cited case. Then the original party can respond completing the 3-ply argument.

CATO extended HYPO with four extra moves:

5. Downplaying the significance of a distinction.
6. Emphasising the significance of a distinction.
7. Citing a favourable case to emphasise strengths.
8. Citing a favourable case to argue that weaknesses are not fatal.

Again the argument is started by one party using the first move to analogise a past case to the problem case. The opponent can then respond to this move using another move and then the original party can respond.

Although HYPO is designed simply to create the arguments whereas CATO is designed to support law students learning how to argue with cases and which move to make at each stage in the argument, CATO can also create its own arguments and explain them. As we will target the basic theory of Bench-Capon and Sartor (2003) rather than the extension designed to allow downplaying and emphasising distinctions, moves 5 and 6 of CATO cannot be used. In any event we would argue that these concern theory evaluation rather than theory construction. Also we will not adopt moves 7 and 8 from CATO at this stage. Arguably these moves also relate to evaluation as they rhetorically strengthen rather than develop the theory. We therefore base AGATHA on the moves found in HYPO, although we use the factor based representation of cases used in CATO rather than dimensions as found in HYPO: again this is because we are using the basic theory of Bench-Capon and Sartor (2003), rather than the extended notion which incorporates dimensions. For an extension to CATE which accommodates dimensions see (Chorley and Bench-Capon, to appear).

AGATHA models the four moves described in HYPO although the distinguish move is expressed as three distinct moves, depending on whether it is the citation of a case, a rule preference or a value preference which is advanced to support the opposing view. The counter example moves have been merged because again the distinction between moves 3 and 4 relates more to evaluation that construction. This gives AGATHA five moves, which we describe in the next section. The idea is that AGATHA will use these moves to simulate a dialogue between the plaintiff and the defendant, constructing the theory as a side effect of the dialogue.

3. Argument moves in AGATHA

Cases are represented as sets of factors and an outcome, which is either plaintiff or defendant. Factors, which represent particular relevant aspects of a case, are represented as a factor name, an outcome favoured by the present of that factor, and a value which is the reason why that factor favours that outcome. See Chorley and Bench-Capon (2003a, b, c) for a full description.

The five moves available in AGATHA are: Analogise Case, Distinguish with Arbitrary Preference, Distinguish with Case, Distinguish Problem, and Counter with Case.

1. *Analogise Case*. This move cites a precedent case which has the outcome the party making the move desires. The factors which are present in both the problem case and the case being cited are sorted into the factors which support that outcome and those factors which support the opposite outcome. A rule preference is made with the supporting factors preferred over the contrary factors. This move follows the method of extracting rules from cases proposed in Prakken and Sartor (1998).
   The first move made has to be *Analogise Case*. *Analogise Case* can also follow the *Distinguish with Arbitrary Preference* move but if using it introduces inconsistencies within the theory, the rule and value preferences that were introduced by the *Distinguish with Arbitrary Preference* move are removed from the theory and then the *Analogise Case* move can introduce new rule and value preferences. It cannot follow the other three moves.
2. *Distinguish with Case*. This move distinguishes a case already cited in the debate and cites a new case which has the different outcome. To distinguish the previously cited case, AGATHA takes all the factors not used in the *Analogise Case* move which support the outcome and adds them to the factors used in the rule preference from the cited case. So, for example, if the previously cited case was a defendant case, AGATHA takes the unused defendant factors from that case and adds them to the used defendant factors. This creates a larger rule containing all the defendant factors from the case which is then preferred over the original plaintiff factors. This gives a more complex rule which can be used to decide the previously cited case but cannot be used to decide the problem case because this case does not contain all the factors contained in the new preferred rule. AGATHA then cites a precedent case with a different outcome from the previously cited case, to give a theory supporting the other side.
   Distinguish with Case can follow the *Analogise Case, Distinguish with Case* and *Counter with Case* moves because these all cite a new case. It cannot follow the *Distinguish with Arbitrary Preference* and *Distinguish Problem* as these do not cite a case.
3. *Distinguish with Arbitrary Preference*. This move distinguishes the previously cited case in the same way as for the *Distinguish with Case* move, but instead of analogising a new case, AGATHA makes an arbitrary preference using the factors from the problem case that are included in the theory constructed by the analogising move and only these factors. If, for example, AGATHA is making a plaintiff move,

the arbitrary preference has the plaintiff factors preferred over the defendant factors, otherwise, for a defendant move, the defendant factors are preferred over the plaintiff factors. The preference is arbitrary because there is no case to support the preference; it just depends on what the party making the move needs to assume to make their case.

It can follow the *Analogise Case*, *Distinguish with Case* and *Counter with Case* moves because they all cite a new case. It cannot follow the *Distinguish with Arbitrary Preference* and *Distinguish Problem* as these do not cite a new case.

4. *Distinguish Problem*. This move distinguishes the problem case instead of the previously cited case. If, for example, AGATHA is making a plaintiff move, it takes all the plaintiff factors from the problem case and conjoins them as the antecedent into a single rule with plaintiff as consequent. The defendant factors from the problem case are similarly conjoined as the antecedent of a single rule with defendant as consequent. Next the value sets comprising the values associated with the factors in the two rules are created and a value preference is created with the value set corresponding to the plaintiff factors being preferred over the value set from the defendant factors. Finally a rule preference is created using this value preference.

   It can follow the *Analogise Case*, *Distinguish with Case* and *Counter with Case* moves because they all cite a new case. It cannot follow the *Distinguish with Arbitrary Preference* and *Distinguish Problem* as these do not cite a new case.

5. *Counter with Case*. This move counters the previously cited case by finding a case which is as-on-point as or more-on-point than the previous case but which was decided for the other side. For an as-on-point counter move, the new case must have the same factors matching the problem case as the previously cited case. The original rule and value preferences which are supported by the previously cited case are replaced with new preferences which are opposite to the original preferences and are supported by the new case. For a more-on-point counter move, the new case must have the same factors matching the problem case as the previously cited case and extra factors which match the problem case but are not present in the previously cited case. The original rule and value preferences supported by the previously cited case are replaced by new preferences which are supported by the new case.

   It can follow the *Analogise Case*, *Distinguish with Case* and *Counter with Case* moves because they all cite a new case. It cannot follow the *Distinguish with Arbitrary Preference* and *Distinguish Problem* as these do not cite a case.

These moves can only be made once to a given theory apart from *Distinguish with Arbitrary Preference* which can be made more than once, and *Counter with Case* and *Distinguish with Case* which can be made once for each of the appropriate cases in the case background. Note also that AGATHA may extend beyond the third ply if moves are available to do so.

The argument moves used in AGATHA use the theory constructors from Bench-Capon and Sartor (2003), Chorley and Bench-Capon (2003a, b, c) to create the underlying theory. When a move is made, a number of theory constructors are applied to extend the current theory. For example, the *Analogise Case* move uses the *Include Case* constructor to include the cited case into the theory, the *Include Factor* constructor to include all the matching factors with the problem case and the *Merge Factors* constructor to merge the plaintiff and defendant factors together. Finally it uses the *Preferences from Case* constructor to include the rule preference which is used to explain the decision for the cited case. Table I shows the Theory Constructors which are used in each move.

4. AGATHA

AGATHA models adversarial dialogue between two agents with each agent taking turns to make a move to produce a theory. As described above, AGATHA has five moves that it can use according to certain preconditions and it applies all possible moves at each point until no more moves can be made.

AGATHA checks which moves can be made by checking the preconditions for each move against the theory at that point in the game tree and, if the preconditions match, it applies the move. Each move that can be applied produces a new theory. When alternative moves are available, new branches are added to the tree of theories being created.

As each move is applied to the theory, the resulting theories are examined and only those which give the same outcome for the problem case as the party making the move are retained. If the move made does not give the desired outcome, the theory is discarded because, even though the move could be applied, it does not help the party making the move, and so does not represent a sensible move.

4.1. MODES IN AGATHA

AGATHA can operate in several modes. In its simplest mode of operation, reported in Chorley and Bench-Capon (2004a, b), it simply generates the complete game tree, the complete theory space. The obvious drawback to this

Table I. Table of theory constructors associated with each move

Move	Theory constructors
Analogise Case	Include Case
	Include Factors
	Merge Factors
	Preferences from case
Distinguish with Case	Include Case
	Include Factors
	Merge Factors
	Preferences from Case
	Remove Rule Preference
Distinguish with Arbitrary Preference	Include Factors
	Merge Factors
	Remove Rule Preference
	Preferences from Case
	Arbitrary Preference
Distinguish Problem	Include Factors
	Merge Factors
	Value Preferences
	Rule Preference from Value Preference
Counter with Case	Include Case
	Include Factors
	Merge Factors
	Preferences from Case
	Remove Rule Preference

brute force approach is that if there is a reasonable number of precedent cases available, there will also be a large number of moves available and the tree rapidly becomes unacceptably large. We have run experiments with 4, 6 and 8 precedent cases (equally split between precedents for the Plaintiff and precedents for the Defendant), and found that highly explanatory theories (comparable with the predictive power of IBP, the currently best performing program) are produced. To produce effective theories, however, requires that "good" precedent cases are chosen, and our objective is to remove such exercises of skill and judgement. We therefore turned to heuristic search methods to enable us to use the complete set of precedent cases while keeping the tree within reasonable limits.

AGATHA's second mode uses a variant of the heuristic search algorithm A* to choose which nodes to expand. The results are reported in Chorley and Bench-Capon (2005). An evaluation of the nodes to drive the heuristic search is given by a program Evaluation of Theories in Law (ETHEL), which assesses a theory according to a number of criteria reflecting both the theory itself (simplicity, and explanatory power) and its value in terms of its position in the game tree (depth and whether it is a leaf node). ETHEL is more fully described in section 6. The results from using A* to guide AGATHA were highly encouraging. In this mode it is possible to generate the theory space using the complete set of cases, and still obtain the level of performance reached using the brute force method with a background composed of carefully selected cases.

A* is not, however, an adversarial search technique. Thus the game played is not competitive, as no account is taken of the responses available to the opponent when a move is made and so no effort is made to block promising counters by the opponent. Law, at least in the US domain we are considering, is adversarial, and so in its third mode AGATHA employs an adversarial search technique, based on standard two-player game techniques for pruning the search.

The following sections describe a series of experiments performed using the different modes of AGATHA. Section 5 shows how the original version of AGATHA performs on two different law domains. Section 6 describes a second program, ETHEL, which is used to evaluate the theories created by AGATHA to show how "good" each theory is for use in heuristic search. The following three sections (7, 8 and 9) describe the first modification made to AGATHA which is the use of the A* search heuristic to guide AGATHA through the theory search space. Sections 10 and 11 describe the second modification to AGATHA which uses the adversarial search heuristic based on alpha beta pruning. Section 12 introduces the final program called ROSALIND which explores the idea of unbalanced information. Finally Section 13 examines the dialogues produced by the winning theories.

5. Experiments in AGATHA

5.1. WILD ANIMALS EXAMPLE

This illustrative example uses the widely discussed wild animal cases used in Bench-Capon and Sartor (2003) and Chorley and Bench-Capon (2004a, b). This small example allows an exhaustive walk through of the operation of AGATHA and permits comparison with the handconstructed theories of Bench-Capon and Sartor (2003).

In all three cases, the plaintiff (P) was chasing wild animals, and the defendant (D) interrupted the chase, preventing P from capturing those animals. The issue to be decided is whether or not P has a legal remedy (a right to be compensated for the loss of the game) against D. In the first case, *Pierson v Post*, P was hunting a fox on open land in the traditional manner using horse and hound, when D killed and carried off the fox. In this case P was held to have no right to the fox because he had gained no possession of it. In the second case, *Keeble v Hickeringill*, P owned a pond and made his living by luring wild ducks there with decoys, shooting them, and selling them for food. Out of malice, D used guns to scare the ducks away from the pond. Here P won. In the third case, *Young v Hitchens*, both parties were commercial fisherman. While P was closing his nets, D sped into the gap, spread his own net and caught the fish. In this case D won.

As analysed in Bench-Capon and Sartor (2003), the cases can be described using four factors and three values. The factors are: the plaintiff did not have possession of the animal (*pNposs*), the plaintiff owned the land (*pLand*), the plaintiff was pursuing his livelihood (*pLiv*) and the defendant was pursuing his livelihood (*dLiv*). These factors are related to three values. The first is intended to reduce litigation (*LLit*); the second to secure enjoyment of property rights (*MSec*) and the last two to promote economically productive activity (*MProd*).

Young is taken as the problem case with *Pierson* and *Keeble* as the set of background cases that AGATHA can use to create the theories. *Keeble* is a plaintiff case and has two factor matches with the problem case. *Pierson* is a defendant case and has one factor matching with the problem case.

Using all the moves defined in AGATHA, AGATHA creates 10 theories which are shown in Figure 1. Figure 1 also shows how the theories relate to each other. The rules, rule preferences and value preferences for the theories produced are shown in Table II.

From Theory 0 (Figure 2) only the *Analogise Case* move can be made. First the defendant move is made by analogising Pierson to the problem case to produce Theory 1 (Table II). Then the plaintiff move is made by analogising Keeble to the problem case to produce Theory 2.

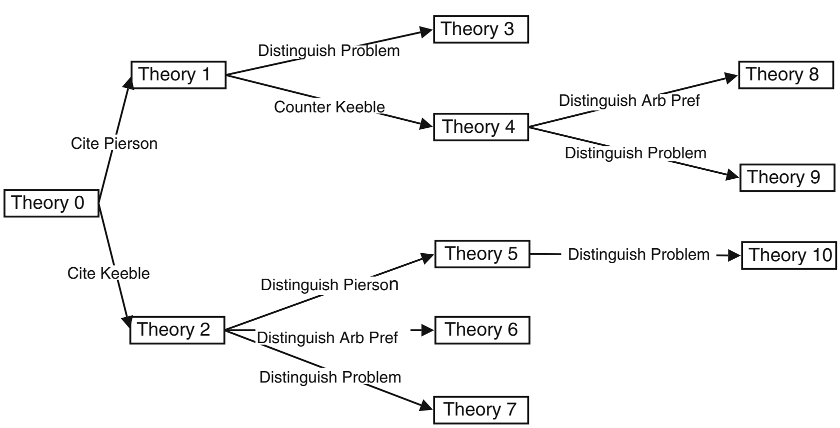

Figure 1. Theories produced.

Table II. The rules, rule preferences and value preferences for the theories

Theory	Rules	Rule preference	Value preference
1	(1) pNposs → D		
2, 3	(1) pNposs → D	(2) > (1)	MProd > LLit
4	(2) pLiv → P		
5	(1) pNposs → D	(4) > (1)	(MProd, MSec) > LLit
	(2) pLiv → P		
	(3) pLand → P		
	(4) (pLiv, pLand) → P		
6, 8	(1) pNposs → D	(4) > (1)	(MProd, MSec) > LLit
	(2) pLiv → P	(1) > (2)	LLit > MProd
	(3) pLand → P		
	(4) (pLiv, pLand) → P		
7, 9	(1) pNposs → D	(2) > (1)	MProd > LLit
	(2) pLiv → P	(4) > (2)	(LLit, MProd) > MProd
	(3) dLiv → D		
	(4) (dLiv, pNposs) → D		
10	(1) pNposs → D	(4) > (1)	(MProd, MSec) > LLit
	(2) pLiv → P	(2) > (1)	MProd > LLit
	(3) pLand → P		
	(4) (pLiv, pLand) → P		

```
Theory Cases :
  <Young, {pLiv, pNposs, dLiv}, D>
Theory Factors :
Theory Rules :
Theory Preferences :
Theory Value Preferences :
```

Figure 2. Theory 0.

From Theory 1 the *Distinguish with Case* and *Distinguish with Arbitrary Preference* moves cannot be made because there are no extra factors that can be used to distinguish Pierson. *Distinguish Problem* can be made to distinguish Young and produce Theory 3. *Counter with Case* can be made because Keeble is more-on-point than Pierson and produces Theory 4. Although, as discussed below, theories 2, 3 and 4 contain the same rules and preferences, the justification of the rules and preferences and the moves available are different for each theory.

From Theory 2 *Distinguish with Case* can be used to distinguish Keeble and cite Pierson to produce Theory 5, *Distinguish with Arbitrary Preference* produces Theory 6 and *Distinguish Problem* produces Theory 7. *Counter with Case* cannot be used because Pierson is less-on-point than Keeble.

From Theory 3 there are no moves that can be made so this line of the dialogue stops. From Theory 4 *Distinguish with Case* and *Counter with Case* cannot be used because there are no more defendant cases to be cited. *Distinguish with Arbitrary Preference* produces Theory 8 and Distinguish Problem produces Theory 9.

From Theory 5 *Distinguish with Case* and *Distinguish with Arbitrary Preference* cannot be used because Pierson has no more factors that could be used to distinguish it. *Distinguish Problem* produces Theory 10. Note that the alternative way of distinguishing the problem, by preferring *MSec* to *LLit* cannot be used because *pLand* is not present in Young, and so this would not produce a pro-plaintiff theory for Young. *Counter with Case* cannot be used as there are no more defendant cases that can be used.

From Theory 6 the only potential move is *Analogise Case*, but this move cannot be used because there are no remaining plaintiff cases.

From Theory 7 there are no moves that can be used.

From Theory 8 the only potential move *Analogise Case*, but again this move cannot be used because there are no remaining plaintiff cases.

From Theory 9 there are no moves that can be used.

From Theory 10 there are no moves that can be used.

The tree is therefore complete.

From an analysis of the preference sections of the theories, it can be seen that several theories have identical preferences, even though these preferences may have different labels because they have been produced using different moves. There are three groups of identical theories and two theories which are different from all the others.

The first group of theories contains Theories 2, 3 and 4. Theories 2 and 4 are identical because Pierson only has one factor and so cannot contribute a rule preference so, for Theory 4 when *Counter with Case* is used, Keeble contributes the same rule preference as *Analogise Case* for Theory 2. Theory 3 is a plaintiff theory and so takes the defendant *pNposs* factor from Theory 1 and adds the plaintiff factor from the Young case description and creates a rule preference of ($pLiv \rightarrow P > pNposs \rightarrow D$) which is the same rule preference which Keeble contributes.

The second group contains Theories 6 and 8. These are identical because their preceding theories are also identical (Theory 2 proceeds Theory 6 and Theory 4 precedes Theory 8) and they are produced by making the same move.

The third and final group contains Theories 7 and 9 and they are identical for the same reasons as the second group.

Theory 5 is a distinct theory. To create Theory 5 from Theory 2, Keeble is distinguished and Pierson is cited but Pierson only has a single factor which is already present in the theory and so does not contribute a rule preference. A pro-defendant outcome is produced, however, because the rule preference of

($\{pLiv, pLand\} \rightarrow P$) over ($pNposs \rightarrow D$), is not applicable to Young, since *pLand* is not present, which allows ($pNposs \rightarrow D$) to fire and give an outcome for the defendant.

This example is also used in Bench-Capon and Sartor (2003) where the theories are produced by hand. In Bench-Capon and Sartor (2003) four theories are produced which correspond to theories 1, 4, 8 and 9. This is because only one branch of the theory space is followed. First Pierson is cited to produce Theory 1 for both examples. Then the opponent uses the *Counter* move with Keeble to give Theory 2 which AGATHA calls Theory 4. Finally the *Distinguish with Arbitrary Preference* move is used to construct Theory 3 and the *Distinguish Problem* move constructs Theory 4. AGATHA calls these Theories 8 and 9 respectively.

These four theories were used to describe a possible process of theory construction using the theory constructors described in Bench-Capon and Sartor (2003). This represents one sensible path through the theory space as constructed by AGATHA. AGATHA also constructs these theories, but also provides alternative theories which may be better (or worse) than those produced by hand.

5.2. CATO EXAMPLE

We have also used AGATHA on our other test domain, US Trade Secrets Misappropriation Law, as modelled in Aleven (1997). The domain is also described in Chorley and Bench-Capon (2003a, b, c). This is a larger domain, containing 32 cases, 26 factors and 5 values.

For this experiment we used the case of *Mason versus Jack Daniels*. The Mason case is described as follows (Aleven 1997).

In 1980, a restaurant owner named Mason developed a combination of Jack Daniel's whiskey, Triple Sec, sweet and sour mix, and 7Up to ease a sore throat. He promoted the drink, dubbed "Lynchburg Lemonade" for his restaurant,"Tony Mason's, Huntsville," served it in Mason jars and sold t-shirts. Mason told the recipe only to his bartenders and instructed them not to reveal the recipe to others. The drink was always mixed out of customer's view. Despite its extreme popularity (the drink comprised about one third of the sales of alcoholic drinks), no other establishment had duplicated the drink, but experts claimed it could easily be duplicated. In 1982, Randle, a sales representative of the distillery, visited Mason's restaurant and drank Lynchburg Lemonade. Mason disclosed part of the recipe to Randle in exchange, Mason claimed, for a promise that Mason and his band would be used in a sales promotion for Jack Daniels. Randle recalled having been under the impression that Mason's recipe was a "secret formula". Randle informed his superior of the recipe and the drink's popularity. A year later, the Distillery began using the recipe to promote the drink in a national sales campaign. Mason did not participate in the promotion or receive other compensation.

Running AGATHA on the case of *Mason versus Jack Daniels* produces the results shown in Table III. With the limited set of background cases of two Plaintiff cases and two Defendant cases,[1] AGATHA produces a tree of depth 7 with 106 nodes. Adding a further two cases[2] gives rise to a theory space with a maximum depth of 8 and 653 nodes. Adding a further two cases[3] produces a tree of depth 11 and 2855 nodes.

As the domain becomes larger, the game tree, and hence the theory space, becomes very much larger. This is entirely to be expected, and as this is what invariably happens to a game tree when we move from a simpler to a more complex game. It means, however, that the exhaustive construction of the theory space will not always be the best strategy for realistically large problems, especially if we want to avoid being selective in the inclusion of cases in the background. We therefore need a means of establishing the worth of theories, so that we can use heuristic search methods.

6. ETHEL – evaluation of theories

ETHEL stands for Evaluation of Theories in Law and evaluates theories using criteria similar to as those proposed in Bench-Capon and Sartor (2003), including explanatory power, simplicity, freedom from arbitrary preferences and the ability to generalise to new cases. ETHEL first analyses the constructed theories to create a table of results reflecting some key metrics of the theory.

6.1. EVALUATION CRITERIA

These metrics will provide the basis for assessment relating to the following five criteria:

Table III. Original AGATHA results

Name	Number of cases		Number of nodes	Tree depth
	Plaintiff	Defendant		
Original1	2	2	106	7
Original2	3	3	653	8
Original3	4	4	2855	11

1. *Simplicity*. Ethel counts the number of rules in the theory, the number of arbitrary rule preferences and the number of rule preferences obtained from value preferences.
2. *Explanatory Power*. Each theory is executed (using a program automatically generated from the theory as described in Chorley and Bench-Capon (2003a, b, c)) with the complete set of background cases and the results analysed. First the total number of cases used is found, and then the number of cases which received the same outcome from the theory as their actual outcome are counted (to give the number of correctly decided cases). Next the cases which received the wrong decision from the theory are counted (giving the incorrectly decided cases) and finally the cases for which the theory could not give an outcome are counted (giving abstentions). This can be used to show how well (or badly) a theory generalises from the cases used in its construction.
3. *Completion Explanatory Power*. The previous criterion executes a program which uses only the factors which are explicitly used in the construction of the theory. This gives a restricted set of factors to be considered when deciding the cases. For the third criterion all the background factors are loaded into the theory, the theory is completed as described in Chorley and Bench-Capon (2003a, b, c) and the program produced from this extended theory is executed on the set of background cases. Again the results are analysed to give the total number of cases, the number of cases which are correctly decided, the number of cases which are incorrectly decided and the number of cases which have no outcome.
4. *Depth*. The theory is given a number corresponding to its depth in the tree.
5. *Leaf Node*. The table indicates if the theory is a leaf node in the tree and hence has no more moves that can be made to reverse its decision.

6.2. EVALUATION PARAMETERS

Ethel now uses this set of metrics to calculate an Evaluation Number for each theory. This is intended to measure how good the theory is and is composed from the above five criteria. For each criterion we provide a way of turning the associated metrics into a number. We also use a number of parameters to control the relative weight given to each criterion. These are to a certain extent pragmatic, justified only by their effect on the evaluation number: those given in this section were our initial choices, and were varied in some of the experiments described below.

1. *Simplicity*. The value for *Simplicity* is composed of three parts; a value based on the number of rules in the theory, a value based on the percentage of the rules that are Arbitrary and a value based on the percentage of the rules that are Value based. The values for the Arbitrary and Value based rules are subtracted from the value for the total number of rules.
   A simpler theory is better than a more complex theory. The simplest theory would only contain one rule preference and this should not be an arbitrary rule preference or a rule preference from value preference.
   If there are no rule preferences then the value is zero: if there is only one rule then the value is 100 and otherwise the value is 100 decreased by a certain percent for each additional rule. In our experiments we used 10% as our discount factor.
   For example, a theory with no rules has a value of zero, a theory with one rule has a value of 100 and a theory with two rules has a value of 90, a theory with three rules 81 and so on.
   The value given by the number of Arbitrary preferences is given by the percentage of the total number of rule preferences which are Arbitrary. If the theory has only one rule and it is Arbitrary then the Arbitrary value is 100 which is subtracted from the total rule value of 100 to give a *Simplicity* value of 0. If the theory has two rules and only one is Arbitrary then the Arbitrary value is 50 and is subtracted from the total rule value of 90 to give a *Simplicity* value of 40.
   The value for the Value-Based rules is calculated in the same way as for the Arbitrary rules but the value used is reduced to two fifths. This is because we consider preferences based on Value preferences to be more principled that those expressed as Arbitrary rule preferences. For example, a theory with one rule which is Value-Based has a value of 40 which is subtracted from the total rule value of 100 to give a *Simplicity* value of 60. If the theory has two rules and only one is value based then the value is 20 and is subtracted from the total rule value of 90 to give a *Simplicity* value of 70.
   Table IV shows some example calculations when the total number of rules is varied as well as the number of Arbitrary and Value-Based rules.
2. *Explanatory Power*. The value for the Explanatory Power is given by the number of correctly decided cases plus half the abstention cases divided by the total number of cases and multiplied by 100 to give a percentage of the total number of cases explained by the theory.
3. *Completion Explanatory Power*. This value is calculated in the same way as for the Explanatory Power.The above three values are summed to give a basic *Evaluation Number* which is based on how well the theory performs in explaining the background cases and its simplicity. We

Table IV. Values calculated for simplicity with different numbers and types of rules

Total rules	Arb. rules	Value-based rules	Total value	Arbitrary value	Value-based value	Simplicity value
0	0	0	0	0	0	0
1	0	0	100	0	0	100
2	0	0	90	0	0	90
3	0	0	81	0	0	81
1	1	0	100	100	0	0
2	1	0	90	50	0	40
3	1	0	81	33.33	0	47.67
1	0	1	100	0	40	60
2	0	1	90	0	20	70
3	0	1	81	0	13.33	67.67
2	1	1	90	50	20	20
3	2	1	81	66.67	13.33	1
3	1	2	81	33.33	26.67	21

now adjust this number according to the position of the theory in the tree.

4. *Depth*. The basic Evaluation Number can be increased by adding a value which represents how deep the theory is in the tree. This encourages AGATHA to explore the search space more deeply. Initially we increased the Evaluation Number by 10% for each additional level greater than level 1.
5. *Leaf Node*. The depth-extended Evaluation Number can be increased again by adding a value which represents whether the theory is a leaf theory, to reflect the fact that this theory cannot be profitably modified by an opponent. If the theory is a leaf theory then the Evaluation Number is increased further, again initially by 10%.

These Evaluation Numbers give a value with which to compare the theories based on how well they explain the background, their structure and their position in the development of the game tree. They can be used to evaluate the nodes in the theory tree, and so guide a heuristic search. We base our first heuristic search on A*, perhaps the best known such algorithm, and described in most standard AI texts (e.g. Winston 1992).

7. AGATHA's version of A* search

A* is not an adversarial search, and so using it in AGATHA makes theory construction a cooperative process as no account is taken of how good a response a move permits, and so involves no notion of blocking the "opponent's" best moves. We will explore the effect of a genuinely adversarial search later in this paper. Standard A* uses two parameters, $f(n)$ which estimates the cost to reach goal from the current node, and $g(n)$ which represents the actual cost of reaching the current node from the initial state. These must be adapted because we do not have any real target: we want only to produce the best possible theory; also it is unimportant how many moves are required to produce it. For $g(n)$, therefore, we use only the cost of the next move from the current theory and do not consider any history of how we reached the node and for $f(n)$ we subtract the *Evaluation Number* for the next Theory from the *Evaluation Number* for the *Ultimate Theory* which is the theory which cannot be improved.

The A* value for each theory is now given by summing the $f(n)$ and $g(n)$ values and only the theories with the lowest A* value are expanded. Before A* starts to work out which move to make, the new theories are checked to ensure that they have a larger *Evaluation Number* than the original theory. This is to ensure that is the new theory represents an improvement.

7.1. $F(N)$ VALUES

To replace the notion of a goal state, we calculate the $f(n)$ value by calculating how similar the next theories are to the best theory possible. This is done by calculating the *Evaluation Number* for the *Ultimate Theory* which consists of one rule and gets all the cases correct for both the *Explanatory Power* section and the *Completion Explanatory Power* section, so that its basic Evaluation Number is maximum. A complete tree with five levels would thus result in an *Ultimate Evaluation Number* of 420.

Now an $f(n)$ value for each theory can be calculated by subtracting the theory *Evaluation Number* from the *Ultimate Evaluation Number*. This means that a "good" theory will have a smaller $f(n)$ value than a "bad" theory because it will be more similar to the *Ultimate Theory* (and have less rules and/or decides more of the background cases correctly) than the "bad" theory and as we want to choose the best theories possible we want to choose the smallest $f(n)$ value.

7.2. $G(N)$ VALUES FOR EACH THEORY MOVE

The $g(n)$ value is given by the cost of making the move to get to the next theory. Each move defined in Section 3 is ranked according to our view of its desirability and associated with a cost. The moves are ranked as follows: *Counter with Case*, *Distinguish with Case*, *Distinguish Problem*, *Distinguish* *with Arbitrary Preference* and finally *Analogise Case*. *Analogise Case* is given the highest value as we wish it to be made only at the beginning, otherwise the dialogue would effectively restart and *Counter with Case* is given the lowest as this is the move that we feel is most desired. The $g(n)$ values for each move are given in Table V. These are intended to reflect our views of which moves would be seen as most powerful by human players.

The A* value for each theory is now given by summing the $f(n)$ and $g(n)$ values and only the theories with the lowest A* value are expanded. This may mean that the ‘‘best’’ theory is not reached becasue that theory may be produced by using a move with a high cost whereas a less good theory which is produced by a move with a low cost will be choosen instead.

8. A* Results

A number of experiments were conducted to explore a series of questions by using various different combinations of parameters. All the experiments described in this section start from the case of *Mason versus Jack Daniels* (discussed in Aleven 1997) as the seed case and all the theories are thus trying to explain the *Mason* case and also the rest of the background cases. These background cases comprise 33 cases taken from various writings on CATO, in particular (Aleven 1997; Brüninghaus and Ashley 2003; Ashley and Brüninghaus, 2003). These cases employ the 26 factors of Aleven (1997), and the five values described in Chorley and Bench-Capon (2003a). Our overall measure of success will be the number of cases that can be explained by a theory. Comparison targets are suggested by Table 1 in Brüninghaus and Ashley (2003). In that paper 10 techniques were tested on 187 cases. The best performer was the algorithm of Ashley and Brüninghaus themselves with 170 right, 15 wrong and one abstention for an accuracy of 91.4%. Next best was Naïve Bayes with an accuracy of 86.5%. No other technique did better than 77.8%. As we are restricted to the 33 cases we have been able to reconstruct from the published literature, 30+ correct classifications would represent a performance comparable to IBP and 28–29 correct classifications

Table V. $g(n)$ values for each move

Move name	$g(n)$
Counter with Case	50
Distinguish with Case	100
Distinguish Problem	150
Distinguish with Arbitrary Preference	200
Analogise Case	250

a performance comparable to Naïve Bayes, and 26 cases a performance better than any other technique considered in Brüninghaus and Ashley (2003).

8.1. HOW DOES A* COMPARE WITH SEARCHING THE COMPLETE SPACE?

Table VI shows the results for the original AGATHA program (Chorley and Bench-Capon 2004a, b), which produced the entire game tree when using a background restricted to 4, 6 and 8 cases. Table VII shows the results obtained by using the A* search on the same limited sets of cases. Table VII also shows how well the best theory produced performs by showing how many of the background cases it gets correct for both the *Explanatory Power* of the theory (only contains factors explicitly stated in the theory) and for the *Completion Explanatory Power* of the theory (all the factors are included in the theory).

A* actually performs worse than the original AGATHA program but this is explained by the fact that A* does not use all the moves available, but chooses the best move which *improves* the theory. A* will only make the move if the next theory is better than the previous theory, whereas AGATHA originally did not impose this condition, and so sometimes there is a theory with a lower Evaluation Number than its previous theory which can subsequently be modified to produce a better following theory. This is not possible within the spirit of adversarial search, since there is no obligation to make a

Table VI. Original AGATHA results

Name	Number of cases		Number of nodes	Tree depth	Best Results	
	Plaintiff	Defendant			Explanatory	Completion
Original1	2	2	106	7	28	28
Original2	3	3	653	8	30	30
Original3	4	4	2855	11	30	31

Table VII. A* results

Name	Number of cases		Number of nodes	Tree depth	Best results	
	Plaintiff	Defendant			Explanatory	Completion
A*1	2	2	8	3	27	28
A*2	3	3	27	6	29	28
A*3	4	4	36	6	29	29

move unless a better theory has been proposed. While we might be able to obtain better results by modifying the parameters in the evaluation function, we felt that the results were close enough to continue our experiments.

The main improvement from using A* is the substantial reduction in the number of theories created (from 106 to 8 theories or 653 to 27 theories) with only a small reduction in the ability of the theory to decide cases correctly. Our hope is that the ability to include more cases from which to select moves resulting from the pruned search space will more than compensate for missing the "best" theory from a limited background. Moreover since the selection of cases depends on the seed case, selection would be difficult to automate since human skill and judgement is needed to find cases appropriate to the seed. Note also that the performance is good in all cases: A* attaining the level of Naïve Bayes, and the exhaustive search the level of IBP when at least six cases are used.

8.2. DOES INCLUDING ALL THE CASES IMPROVE THE THEORIES?

It is not desirable for the cases to be selected by a human user: we want AGATHA itself to select the best cases to cite from the whole background set. Using all the cases, however, is not viable without using a search heuristic because the search space is too large.

In fact, as can be seen from Table VIII, the experiments made with all the cases available but which select only Most on Point when using A* perform worse than the experiments with selected cases. This might be seen as vindicating our selection of cases for AGATHA, but as will be discussed later, also suggests that the restriction to Most on Point cases is not desirable.

8.3. IS THE DEPTH CONTRIBUTION IMPORTANT?

To see how important considering depth is we ran several experiments using the Most on Point cases and varied the contribution of depth to the Evaluation

Table VIII. A* results using the Most on Point cases

Name	Depth Contribution	Number of nodes	Tree depth	Best results	
				Explanatory	Completion
mop1	0	26	4	25	26
mop2	10	56	5	24	23
mop3	20	240	13	25	26
mop4	50	638	13	27	26

Number from 0 to 50%. For Tables VI and VII the contribution of depth was fixed at 10% but for Table VIII it ranges from 0 to 50%.

When comparing the experiments which only select Most on Point cases (mop1 to mop4 in Table VIII) we find there is no improvement in the number of cases decided correctly after the depth factor reaches a value of 25% of the evaluation power.

Increasing the contribution of depth means that it is easier for AGATHA to use the *Distinguish with Arbitrary Preference* move because the depth value helps to counteract the large $g(n)$ value.

Although increasing the depth contribution means that more moves could be made to the theories and so allow greater exploration of the space it appears that this effects little real improvement in the quality of the theories.

An additional problem with having a deeper tree means that the theories may become over fitted to the cases used and hence do not generalise very well.

8.4. IS IT BETTER TO USE THE MOST ON POINT CASES OR ALL THE CASES?

Using only Most on Point cases means that AGATHA can only use some of the cases from the background as determined by the seed case. This limits AGATHA to a small subset of the background cases. Although this means that intuitively only the most pertinent cases are used, these cases are only most pertinent for the seed case and these cases may provide insufficient information to decide the remaining cases from the case background. As part of the evaluation of the theory depends on how well the theory can generalise to other cases which may have little in common with the seed case, this limitation of cases may be undesirable. Therefore AGATHA was modified to use all the background cases to create the theories, giving the results (both ignoring and including a depth factor) shown in Table IX.

HYPO and CATO use the Most on Point cases because they are concerned with only one case and creating an argument to explain this one case. Using Most on Point cases can limit the moves available to the opponent and prevent counter attacks. AGATHA is different in that it is trying to explain all the background cases not just one case and so it may be that a less on

Table IX. A* results using A* results using all cases

Project name	Depth	Number of nodes	Tree depth	Best results	
				Explanatory	Completion
allCases1	0	93	5	28	29
allCases2	10	2427	16	31	30

point case produces a better theory which can explain more of the background cases.

When AGATHA is able to choose from all of the cases in the background set, the theories improve on the Most on Point theories. Even when the depth is set to 0 (allCases1) the All Cases theories perform better than the Most on Point cases because they get 28 out of 33 cases correct for the Explanatory Power and 29 out of 33 cases correct for Completion Explanatory Power compared with 25 out of 33 cases correct for the Explanatory Power and 26 cases out of 33 correct for Completion Explanatory Power when restricted to Most on Point cases.

Even when comparing All Cases with a depth of 0 with the Most on Point versions with a depth power up to 50 the theories using all the background cases are the best as they have better results while generating a smaller tree.

When the depth value is raised to 10 (allCases2) the results again improve getting 31 out of 33 cases correct for the Explanatory Power and 30 out of 33 cases correct for Completion Explanatory Power compared with 28 out of 33 cases correct for the Explanatory Power and 29 cases out of 33 correct for Completion Explanatory Power, although size of the tree rises from 93 nodes to 2427 nodes. This level of performance is also better than the complete search on 6 cases, and, with respect to Explanatory Power, better than complete search with eight cases. The performance here is comparable to IBP, and we regard the increase in the search space, which remains less than the complete search with a background of only eight cases, as acceptable.

8.5. IS THE COST OF THE MOVES CORRECT?

In all the previous experiments the different moves are ranked with a different cost for each move, as shown in Table V. However we wanted to test the hypothesis that the *Distinguish with Case* move is at least as desirable as the *Counter with Case* move. Table X shows the results when the moves *Counter with Case* and *Distinguish with Case* are given the same $g(n)$ value of 50.

With equal weights the *Distinguish with Case* move is made more often, and the performance is broadly similar. When comparing all- Cases1 and sameWeight1, not penalising the *Distinguish with Case* move means that the

Table X. A* results counter and distinguish same weight using all cases

Project name	Depth	Number of nodes	Tree depth	Best results	
				Explanatory	Completion
sameWeight1	0	116	4	29	30
sameWeight2	10	572	7	29	30

total number of theories increases by 23 but the depth of the tree decreases by one level and sameWeight1 gets more cases correct. However when comparing allCases2 and sameWeight2, not penalising the *Distinguish with Case* move means that the number of theories decreases significantly and the depth of the tree halves. However, the larger number of theories does mean that allCases2, where *Counter with Case* is preferred, gets more cases correct in the Explanatory section and hence has a larger Evaluation Number.

8.6. SUMMARY OF RESULTS ON MASON

From these results we conclude that using all the background cases is preferable to using only the Most on Point cases and that the contribution of depth is of some importance: 10% giving better results. Particularly interesting is the improvement given by using cases which are not the most on point. On pointedness is important in both HYPO and CATO, and using such cases has a tactical point in that they are the least open to distinction. On the other hand, using portions of precedents has also long had its advocates (e.g. Branting 1991). Deciding a case often involves considering a number of sub issues and it may well be that a precedent is very relevant for one of these sub issues, although otherwise very dissimilar from the case under consideration. To include such cases starting from a given seed, therefore, we need to go beyond the set of cases most on point to the seed case. Whether on pointedness becomes more useful with adversarial search is something we shall consider later in the paper. Issue based selection of cases is also a feature of IBP (Brüninghaus and Ashley 2003). Using different weights for the *Distinguish with Case* and *Counter* moves and preferring the *Counter* move gives better results consistent with our original view on move costs.

9. Use of other cases as the problem case

In all of the experiments described in Section 8 the case of *Mason versus Jack Daniels* was used as the seed case. In this next set of experiments we wanted to explore the use of other cases as the seed case.

For all of the experiments in this section AGATHA used all the cases, a depth factor of 10% and different weights for the *Distinguish with Case* and *Counter* moves.

We chose a range of different cases to test various classes of case: these cases are shown in Table XI and the results of each experiment are shown in Table XII.

Table XI. Cases and the types and numbers of factors which describe them

Case	Number of factors	Plaintiff	Defendant	Case outcome
Sandlin	5	0	5	D
Ferranti	5	1	4	D
Reinforced	6	5	1	P
Boeing	7	5	2	P
Technicon	7	4	3	P
CMI	7	2	5	D
College Watercolor	3	2	1	P
Sheets	3	1	2	D

Table XII. Results when using different problem cases different weights

Project name	No. of nodes	Tree depth	Best results	
			Explanatory	Completion
Sandlin	10	1	11	24
Ferranti	20	2	8	24
Reinforced	60	3	22	29
Boeing	95	3	29	29
Technicon	2193	17	31	31
CMI	220	7	29	31
College Watercolor	98	5	24	24
Sheets	70	3	26	29

The first experiment used the case of *Sandlin* to see what happens when the problem case only has one type of factor as *Sandlin* only has defendant factors.

Because there are only defendant factors, only the defendant player can make a move, which is why the theory tree only has 10 theories and a depth of 1. The Plaintiff player cannot make a move because there are no plaintiff factors to use.

The theories constructed do not perform very well as they only get 11 cases correct out of 33 for the Explanatory Power and 24 correct out of 33 for the Completion Explanatory Power. This shows that AGATHA can only perform effectively when there are factors from both sides present in the seed case. When there are only factors of one type then no rule preferences can be made so the effectiveness of the theory depends only on the alpha-numerical sorting of the rules during execution and evaluation.

There are no cases in the background with only Plaintiff factors so we could not perform the reciprocal experiment. Instead we chose a Defendant case with only one Plaintiff factor and a Plaintiff Case with one defendant factor.

For the Defendant case we used *Ferranti*, which has one Plaintiff factor and four Defendant factors. The theory tree has more theories and goes to an extra level. However it performs much worse compared to using *Sandlin* as the seed case as it only gets eight cases correct out of 33 for the Explanatory Power. This low number arises from the very high number of abstentions each theory makes.

We used *Reinforced-Moulding* as the reciprocal Plaintiff case as it has five Plaintiff factors and only one Defendant factor. The theory tree again has more theories and reaches a depth of 3. It also gets much better results with 22 cases correct out of 33 for the Explanatory Power and 29 correct out of 33 for the Completion Explanatory Power.

These three experiments show that having factors of both types present in the problem case means that AGATHA can use the Argument Moves effectively to improve the theories. If the seed case contains few factors, completion of the theory seems essential, as otherwise there are too many abstentions.

To explore this point further we investigated whether the number of factors in the case description is important. To show this we split the experiment in two and chose cases with the most number of factors and cases with the smallest number of factors.

For the large cases we chose *Boeing*, which is a Plaintiff case with five Plaintiff factors and two Defendant factors, *Technicon*, which is a Plaintiff case with four Plaintiff factors and three Defendant factors, and finally *CMI*, which is a Defendant case with two Plaintiff factors and five Defendant factors. We chose three cases because we wanted to compare a very Plaintiff case, a very Defendant case and a balanced case.

When the experiments were run the larger cases improved over the first set of experiments. Of the three experiments, using *Technicon* as the problem case performed the best as it got 31 cases correct out of 33 for both the Explanatory Power and the Completion Explanatory Power, the best combined result obtained during our series of experiments (although the completed version of *CMI* was one case better). However the theory tree is quite large, containing over 2000 theories and reaching a depth of 17. Other than *CMI*, completing the theory by including all the factors gives no improvement.

This suggests that a more balanced problem case will produce a larger theory tree and obtain better results than if the case is biased towards one of the parties, especially with respect to the uncompleted theory.

For the experiments with the smallest cases, we chose *College Watercolor*, which is a Plaintiff case with two Plaintiff factors and one Defendant factor and *Sheets*, which is a Defendant Case with one Plaintiff factor and two Defendant factors. When the experiments were run, *Sheets* performed better because it obtained 26 cases correct out of 33 for the Explanatory Power and 29 correct out of 33 for the Completion Explanatory Power compared to 24 correct for both the Explanatory Power and the Completion Explanatory Power even though the theory tree for *Sheets* has fewer theories and two fewer levels. When comparing the small cases with the large cases, the larger cases perform better, and, for small cases, completion gives improvement.

When comparing all of these experiments, the experiments with very biased problem cases perform worst whereas the experiments with wellbalanced problem cases perform best. For the question of the size of the cases, the problem cases with the most factors perform better than those with fewer factors.

These experiments also show that the tree must go to at least the third level to get good results. This seems to correspond to the 3-ply arguments of HYPO and CATO.

Overall we would conclude that the best way to generate a theory automatically would be to select as seed the most balanced background case with the most factors, and use A* with all cases and a depth factor of 10%. This technique, represented by the entry for *Technicon* in Table XII, produces a theory which gives performance at a level similar to that of IBP.

10. Adversarial search method

In this mode AGATHA is modelling an adversarial dialogue. Here the two agents both want to win and prevent the other from winning. This is different from the cooperative search heuristic mode where the two agents were cooperating with each other to produce the best theory and it was not important who won.

The two agents have to pick their way through the theory space by looking ahead to find the best theory for them and try to aim towards it. But they want to prevent their opponent from making a better theory and so they may not choose the path to the best theory if their opponent can create an even better theory and win. We restrict look ahead for 3 moves.

In modelling this dialogue, AGATHA applies a 3-ply method on each theory that has been created. AGATHA first applies all the moves to the theory that are possible to create a group of 1-ply theories. These 1-ply theories represent all the moves that the player can make in the current situation and AGATHA has to choose which of them to play.

AGATHA takes the first 1-ply theory and expands it by only one move. This move represents how the opponent could respond to the theory. AGATHA then expands this 2-ply theory by one move to give the 3-ply theory which the current player could make in response to the opponents theory. AGATHA then assesses the 3-ply theory using the ETHEL program and stores the 3-ply theory.

AGATHA then takes the next 1-ply theory and expands it to the 3-ply theory. Its Evaluation Number is calculated by ETHEL and compared to the first Evaluation Number. If the new theory is better then the 3-ply theory is stored otherwise it is discarded.

AGATHA continues in this way until there are no more 1-ply theories left. It then finds the best 3-ply theory using ETHEL and the grandparent 1-ply theory of this theory will represent the best move for the player to make. Ties are broken using the 2-ply theories: the worst of these is chosen, to restrict the opponent's opportunities.

When the search tree is exhausted and the dialogue ends, AGATHA finds the best Plaintiff theory and the best Defendant theory for each branch in the pruned tree.

11. Results of adversarial search

11.1. COMPARISON WITH COMPLETE SPACE AND A* RESULTS

Table XIII shows the results when the adversarial version of AGATHA is restricted to the 4, 6 and 8 cases that were used earlier in the original and A* versions of AGATHA. The Adversarial version of AGATHA explores more theory nodes and expands the tree to a greater depth than the A* version but expands fewer theory nodes than the original version. The Adversarial version obtains better results than A* and the results are almost as good as the original version.

11.2. USING ALL CASES

Table XIV shows the results obtained when AGATHA is restricted to only using the Most on Point cases from the background (labelled adverMOP) and when AGATHA is allowed to use all the cases in the background.

Table XIII. Adversarial results

Name	Cases		Nodes	Depth	Plaintiff results		Defendant results	
	P	D			Explanatory	Completion	Explanatory	Completion
adver1	2	2	19	5	27	28	27	26
adver2	3	3	39	7	30	30	29	28
adver3	4	4	60	9	29	30	30	30

Table XIV. Adversarial Results

Name	Nodes	Depth	Plaintiff results		Defendant results	
			Explanatory	Completion	Explanatory	Completion
adverMOP	98	12	27	28	27	26
adverAll	779	39	30	31	30	30

AGATHA obtains better results when it is allowed to use all the background cases instead of just the Most on Point cases. This is the same result as for the A* version and confirms the A* finding that to obtain a good explanatory theory AGATHA needs to use all of the background cases in whichever version is being used.

When both players can use the full background AGATHA produces a deeper tree than when using A* search but visits far fewer nodes, e.g. 779 rather than 2427 when using all the cases, and obtains a similar quality of results. So when both sides are fully informed adversarial search improves efficiency without degradation of performance. Overall therefore, there seem to be advantages in using Adversarial rather than Co-operative search.

12. ROSALIND

Using AGATHA, both players have access to the complete set of precedent cases. To explore the effect of different information being available to the two parties we produced ROSALIND which enables the Plaintiff and Defendant each to have their own sets of cases.

There is little effect to the results if one player has all precedents and the other only the precedents favouring its own side. But if the information available to one side is inferior (for example, in terms of the number of factors in its precedent cases) to that of the other the final theories are significantly worse: the better equipped player does not have to produce a very high quality theory to win the game.

Table XV shows all the cases from the case background sorted by size into five groups. The Defendant cases are labelled to show how they are spread across the five groups.

In our experiments using unbalanced information we used three cases, all taken from the group with five factors. Mason was chosen as a balanced case, Bryce as a strongly pro-Plaintiff case and Ferranti as a strongly pro-Defendant case. For the first experiment (labelled P1) the Plaintiff is given all of the cases from the group with seven factors and the Defendant is given is given the cases from the group with only three factors. The second experiment (labelled P2) has the Plaintiff using the two groups with the largest number of

Table XV. Cases sorted by size depending on the number of factors describing the cases

Three factors	Four factors	Five factors	Six factors	Seven factors
Arco (d)	Ecologix (d)	Bryce	Digital development	Boeing
College Watercolor	Forrest	Den-Tal-Ez	FMC	CMI (d)
Emery	Goldberg	Ferranti (d)	MBL (d)	KG
Lewis	Mineral Deposits	Laser	Mineral Deposits Two	National Rejectors (d)
National Instrument	Yokana (d)	Robinson (d)	Reinforced	Technicon
Sheets (d)		Sandlin (d)	Scientology (d)	Televation
		Space Aero	Valco-Cincinnati	
		Trandes		

factors (the cases with six and seven factors) and the Defendant receives the two groups with the smallest number of factors. The third and fourth experiments (labelled D1 and D2 respectively) are the reverse with the Defendant receiving the larger cases and the Plaintiff the smaller cases.

The results when Mason is the seed case are shown in Table XVI. Mason gets the best results when compared with the other seed cases because it is a well balanced case. The Plaintiff usually gets the best results even when it is disadvantaged with the smaller cases. The results improve when more cases are used but it actually gets the best results when the Plaintiff player is disadvantaged and the Defendant player has the larger cases. This may because the player with the case which won in practice is pushed harder when the opponent has better cases, and so is driven to produce a better theory to win.

The results when Bryce is used as the seed case are shown in Table XVII. The Plaintiff player always gets the best results even when it is disadvantaged. The Plaintiff player starts all the dialogues (except for bryceP2 where both players can start the dialogue) and the only moves that the Defendant player can make are Distinguish with Arbitrary Preference and Problem Distinguish.

Table XVI. Results with Mason as seed when choice of case is biased by size to alternately Plaintiff and Defendant

Name	Case size		Nodes	Depth	P results		D results	
	P	D			Explanatory	Completion	Explanatory	Completion
masonP1	7	3	21	8	23	23	21	25
masonP2	6,7	3,4	148	19	27	28	26	26
masonD1	3	7	30	6	27	29	28	28
masonD2	3,4	6,7	41	8	27	29	27	27

With all cases available to both 30-31 cases are explained.

Table XVII. Results with Bryce as seed when choice of case is biased by size to alternately Plaintiff and Defendant

Name	Case size		Nodes	Depth	P results		D results	
	P	D			Explanatory	Completion	Explanatory	Completion
bryceP1	7	3	5	2	12	15	10	11
bryceP2	6,7	3,4	43	18	24	28	8	12
bryceD1	3	7	3	2	19	11	8	21
bryceD2	3,4	6,7	9	3	19	28	19	28

This is because the only Defendant factor in Bryce is F1 and the only Defendant case with this factor is Ecologix in the 4 factor group and the dialogue can only proceed past a depth of 3 when the Defendant player can use this case.

The results when Ferranti is used as the seed case are shown in Table -XVIII. Ferranti produces relatively poor quality theories because the only Plaintiff factor in Ferranti does not appear in any of the other Plaintiff cases and so none of them can be used. The tree only has a depth of 2 because the Defendant player Analogises a Defendant case and then the Plaintiff player responds with the Problem Distinguish move. The Defendant player always gets better results even when it is disadvantaged by having the smaller cases.

13. Argumentation dialogues

The results summarised in the previous sections are highly encouraging with respect to the aim of producing an automated way of constructing theories capable of effectively explaining the precedent cases. Given a balanced seed case and full information available to both parties, highly explanatory theories can be produced, and using adversarial search improves efficiency without

Table XVIII. Results with Ferranti as seed when choice of case is biased by size to alternately Plaintiff and Defendant

Name	Case size		Nodes	Depth	P results		D results	
	P	D			Explanatory	Completion	Explanatory	Completion
ferrantiP1	7	3	4	2	8	14	8	24
ferrantiP2	6,7	3,4	8	2	8	14	8	24
ferrantiD1	3	7	4	2	8	14	8	24
ferrantiD2	3,4	6,7	8	2	8	14	8	24

degrading performance. In this section we will consider our second objective, which was to produce such theories through a plausible argumentation dialogue.

We compare dialogues based on the three seed cases mentioned above each of which contains five factors: Mason, chosen because it is balanced, Bryce, chosen as a strongly pro-Plaintiff case and Ferranti, chosen as a strongly pro-Defendant case. We begin by showing the dialogues produced by AGATHA using A*.

Because some of the dialogues will be described giving details of the rule preferences the descriptive names of the factors are included in Table XIX.

13.1. A* DIALOGUES

Table XX shows the winning dialogues for A* search. The first column shows the dialogue when using Mason as the seed case. The dialogue produces an excellent theory which obtains results of 31 cases correct out of 33 cases when using the factors produced in constructing the theory and 30 cases correct out of 33 cases for the full set of available factors.

The dialogue for Mason is shown in Table XXI. It starts with the Defendant player using the *Analogise* move with the Sandlin case, which contains both the pro-Defendant factors in Mason. This gives a complex rule of $\{(F1, F16) \rightarrow D\}$ but as there are no matching Plaintiff factors in the Sandlin case there is no rule preference produced. The Plaintiff player responds by using the *Counter* move with the modified Mineral Deposits

Table XIX. Factors in CATO (NB: There is not F9 in [Alv])

Pro plaintiff factors	Pro defendant factors
F2 Bribe employee	F1 Disclosure in negotiations
F4 Agreed not to disclose	F3 Employee sole developer
F6 Security measures	F5 Agreement not specific
F7 Brought tools	F10 Secrets disclosed outsiders
F8 Competitive advantage	F11 Vertical knowledge
F12 Outsider disclosures restricted	F16 Info reverse engineerable
F13 Non-competition Agreement	F17 Info independently generated
F14 Restricted material used	F19 No security measures
F15 Unique product	F20 Info known to competitors
F18 Identical products	F23 Waiver of confidentiality
F21 Knew info confidential	F24 Info obtainable elsewhere
F22 Invasive techniques	F25 Info reverse engineered
F26 Deception	F27 Disclosure in public forum

Table XX. A* search dialogues

A* Mason	A* Bryce	A* Ferranti
D Cite Sandlin	D Cite Robinson	D Cite CMI
P Counter with Mineral Deposits Two	P Distinguish with Lewis	
D Distinguish with Yokana		
P Counter with Technicon		
D Distinguish with Arco		
P Distinguish with National Instrument		
D Distinguish with National Rejectors		
P Distinguish with Goldberg		
D Counter with Ecologix		
P Counter with Lewis		
D Distinguish with CMI		

Table XXI. Explanation of A* Mason dialogue

Theory move	Factors in cited case	Rules produced
D Cite Sandlin	$F1,F10,F16,F19,F27$	$(F1,F16) > ()$
P Counter with Mineral Deposits Two	$F1,F4,F16,F18,F21,F25$	$(F21) > (F1,F16)$
D Distinguish with Yokana	$F7, F10, F16, F27$	$(F4,F18,F21) > (F1,F16)$ $(F16) > ()$
P Counter with Technicon	$F6,F10,F12,F14,$ $F16,F21,F25$	$(F6,F21) > (F16)$
D Distinguish with Arco	$F10,F16,F20$	$(F6,F12,F14,F21) > (F16)$ $(F16) > ()$
P Distinguish with National Instrument	$F1,F18,F21$	$(F10,F16,F20) > ()$ $(F21) > (F1)$
D Distinguish with National Rejectors	$F7,F10,F15,F16,$ $F18,F19,F27$	$(F18,F21) > (F1)$ $(F16) > (F15)$
P Distinguish with Goldberg	$F1,F10,F21,F27$	$(F10,F16,F19,F27) > (F15)$ $(F21) > (F1)$
D Counter with Ecologix	$F1,F19,F21,F23$	$(F1) > (F21)$
P Counter with Lewis	$F1,F8,F21$	$(F21) > (F1)$
D Distinguish with CMI	$F4,F6,F10,F16,F17,F20,F27$	$(F8,F21) > (F1)$ $(F16) > (F6)$

case,[4] which has both the pro-Defendant factors and one of the pro-Plaintiff factors in Mason. This move creates a rule preference of $(\{(F21) \rightarrow P\} > \{(F1, F16) \rightarrow D\})$ which prefers the new rule of $\{(F21) \rightarrow P\}$ to the initial rule of $\{(F1, F16) \rightarrow D\}$.

The Defendant then responds by *Distinguishing* the Mineral Deposits Two case, which has *F4-Agreed Not to Disclose* and *F18-Identical Products*, pro-Plaintiff factors present in Mineral Deposits Two but not Mason and then *Analogising* using the Yokana case. The distinguish part modifies the previous rule preference by including all the remaining Plaintiff factors from Mineral Deposits Two to give a rule preference of $(\{(F4, F18, F21) \rightarrow P\} > \{(F1, F16) \rightarrow D\})$. The analogise part does not produce a rule preference as there is only one Defendant factor in Yokana which matches the factors from the seed case (*F16-Info Reverse Engineerable*). The Plaintiff then *Counters* with the Technicon case because Technicon has the Defendant factor from Yokana as well as two Plaintiff factors that Yokana does not have and this produces a rule preference of $(\{(F6, F21) \rightarrow P\} > \{((F16) \rightarrow D\})$.

The Defendant then responds by *Distinguishing* the Technicon case to modify the previous rule using strengths for the Plaintiff not in Mason, and then *Analogising* using the Arco case, another case found for the Defendant with *F16-Info Reverse Engineerable*. The Plaintiff then Distinguishes Arco and Analogises with National Instrument, again using *F21-Knew Info Confidential*. The attempt to establish one of the Defendant factors as superior to *F21-Knew Info Confidential* continues until the Defendant *Distinguishes* Lewis and *Analogises* with the Defendant case CMI. The Plaintiff cannot respond and so the Defendant player wins (contrary to how the case was actually decided). The dialogue shows that the major point for the Plaintiff is that the Defendant knew that the information was given in confidence and that the major point for the Defendant is that the information could have been reverse engineered. The cases are deployed in an effort to determine the context in which these points can be decisive for one side or the other. At the end of the dialogue the theory suggests that if the information could have been reverse engineered, the Plaintiff needs to show that the Defendant gained some competitive advantage by breaking the confidence (*F8-Competitive Advantage*). This seems a plausible contention. *F8- Competitive Advantage* was not present in Mason, and although Mason won, the paltry damages awarded indicate the perceived value of the secret. We would argue that, while this dialogue may not correspond to a dialogue that lawyers would normally produce – the one point nature of our dialogue moves as against the multi-point moves made in actual submissions militates against this – the cases deployed make plausible points, and identified the key point of contention.

The Bryce dialogue is shown in the second column of Table XX. This dialogue produces a theory which obtains 27 cases out of 33 for selected factors and 28 cases out of 33 for the full set of factors.

The Defendant starts the dialogue by *Analogising* with the Robinson case to produce a rule preference of $(\{(F1) \rightarrow D\} > \{((F18) \rightarrow P\})$. The Plaintiff player then responds by *Distinguishing* Robinson which modifies the previous rule preference to $(\{(F1, F10, F19) \rightarrow D\} > \{((F18) \rightarrow P\})$ and then

Analogising with the Lewis case to introduce a new rule preference of $(\{(F21) \rightarrow P\} > \{((F1) \rightarrow D\})$. The Defendant cannot respond and so the Plaintiff player wins.

The Ferranti dialogue only gets nine cases out of 33 for selected factors and 24 cases out of 33 for the full set of factors and only has one move. The Defendant starts the dialogue by *Analogising* with the CMI case which does not produce a rule preference. There may be several moves that the Plaintiff player could make but they do not improve the theory so the dialogue stops and the Defendant player wins.

Because Mason is a well balanced case and contains factors that are present in many of the background cases AGATHA can use many more moves to create longer dialogues which continually refine the theory to produce a very good theory which can decide a large proportion of the background cases correctly.

Bryce also contains factors which are present in many of the background cases but it is very unbalanced as it only contains one Defendant factor. This limits the number of moves that the Defendant can make and so the dialogue is small and there are not many refinements that can be made by the Defendant to the theory to improve it. In consequence the Plaintiff is not required to refine his theory beyond what is necessary to meet these limited objections.

Ferranti is very unbalanced to the Defendant and several of the Defendant factors are present in most of the background cases. However the single Plaintiff factor is not present in any of the background cases so the Plaintiff player is severely limited in the moves that it can make. This means that the dialogue is very short with little or no refinement of the theory.

13.2. ADVERSARIAL DIALOGUES IN AGATHA

When using adversarial search in AGATHA, we get the dialogues shown in Tables XXII and XXIII. The theories obtained for the Plaintiff and Defendant players are identical except for the last move which is shown in bold type. This move represents the final move that can be made in the dialogue and so ends the dialogue. So for the Mason dialogue the Plaintiff theory is complete when the Plaintiff *Analogises* with the Goldberg case but the Defendant extends this theory by using the *Problem Distinguish* move to create the Defendant theory.

All the dialogues are longer than those obtained using A* search and the ones for Mason and Bryce are much longer. This is because the *Distinguish with Arbitrary Preference* is used much more and especially by the Defendant player. We might expect this: a co-operative opponent engaged in seeking the best solution will ground his objections in cases, but in so doing invites more powerful responses. When a player is also trying to give as few opportunities

Table XXII. Adversarial search dialogues – part one

Mason	Bryce	Ferranti
D Cite Ecologix	P Cite Valco-Cincinnati	D Cite Sheets
P Distinguish with Emery	D Distinguish with Sandlin	**P Problem distinguish**
D Distinguish with Arb Pref	P Distinguish with Boeing	
P Cite FMC	D Distinguish with Arb Pref	
D Distinguish with CMI	P Cite SpaceAero	
P Distinguish with Boeing	D Distinguish with Arb Pref	
D Distinguish with Sandlin	P Cite Digital Development	
P Distinguish with Mineral Deposits Two	D Distinguish with Arb Pref	
	P Cite Den-Tal-EZ	
D Distinguish with Arco	D Distinguish with Arb Pref	
P Distinguish with Bryce	P Cite Laser	
D Distinguish with Yokana	D Distinguish with Arb Pref	
P Distinguish with Laser	P Cite FMC	
D Distinguish with Robinson	D Distinguish with Arb Pref	
P Distinguish with Den-Tal-EZ	P Cite Reinforced	
	D Distinguish with Arb Pref	
D Distinguish with National Rejectors	P Cite Lewis	
	D Distinguish with Arb Pref	
P Distinguish with Valco-Cincinnati	P Cite Forrest	
	D Distinguish with Arb Pref	
D Distinguish with Arb Pref	P Cite Trandes	
P Cite College Watercolor	D Distinguish with Arb Pref	
D Distinguish with Arb Pref	P Cite Mason	
P Cite SpaceAero	D Distinguish with Arb Pref	
D Distinguish with Arb Pref	**P Cite Emery**	
P Cite Digital Development		
D Distinguish with Arb Pref		
P Cite KG		
D Distinguish with Arb Pref		
P Cite Lewis		
D Distinguish with Arb Pref		
P Cite National Instrument		
D Distinguish with Arb Pref		
P Cite Reinforced		
D Distinguish with Arb Pref		
P Cite Technicon		
D Distinguish with Arb Pref		

Table XXIII. Adversarial search dialogues – part two

Mason	Bryce	Ferranti
P Cite Televation		
D Distinguish with Arb Pref		
P Cite Trandes		
D Distinguish with Arb Pref		
P Cite Goldberg		
D Problem distinguish		

to his opponent as possible, however, hypothetical objections, not grounded in any cases, can be used to obstruct the deployment of these powerful cases. In Mason the second Defendant move is of this sort as are many of the later moves. For the Bryce dialogue all the moves that the Defendant player makes are *Distinguish with Arbitrary Preference* except the first move where it *Distinguishes* the Valco-Cincinnati case and *Analogises* with the Sandlin case.

For the Mason dialogue the Plaintiff theory gets better results and so wins. The dialogue is comparable in Explanatory Power to A* even though the dialogues are very different. However the Adversarial dialogue contains all of the cases used in the A* dialogue and most of the same moves, although the use of arbitrary preferences changes the order, and forces the introduction of extra cases. Note again, however, that the tree in the adversarial search is narrower, and so fewer nodes are considered in selecting the moves.

The Mason dialogue for Adversarial search is shown in Tables XXIV, XXV and XXVI. The Defendant starts the dialogue by *Analogising* the Ecologix case, which is a good case to choose because it is an example of where the Plaintiff lost despite having *F21-Knew Info Confidential* the main plaintiff point according to the A* dialogue. This results in the rule preference of $(\{(F1) \rightarrow D\} > \{(F21) \rightarrow P\})$. The Plaintiff then responds by *Distinguishing* Ecologix by including the Defendant factors which are present in Ecologix and not in Mason which modifies the rule preference to $(\{(F1, F19, F23) \rightarrow D\} > \{(F21) \rightarrow P\})$ and *Analogises* the Emery case. However because Emery does not have any Defendant factors matching with Mason there is no rule preference produced.

The Defendant then *Distinguishes* the Emery case because there is an extra Plaintiff factor present (*F18-Identical Products*) which could have helped to decide Emery which is not present in Mason. For the second part of the move the Defendant states an *Arbitrary Rule Preference* which is actually the same rule preference that was produced by the first move.

The Plaintiff's response to this move is to effectively restart the dialogue by *Analogising* with the FMC case, although this case only has a single Plaintiff factor matching with Mason and so does not produce a rule preference.

Table XXIV. Explanation of adversarial Mason dialogue – part one

Theory move	Factors in cited case	Rules produced
D Cite Ecologix	*F*1,*F*19,*F*21,*F*23	(*F*1) > (*F*21)
P Distinguish with Emery	*F*10,*F*18,*F*21	(*F*1,*F*19,*F*23) > (*F*21) (*F*21) > ()
D Distinguish with Arb Pref		(*F*18,*F*21) > () (*F*1) > (*F*21)
P Cite FMC	*F*4,*F*6,*F*7,*F*10,*F*11,*F*12	(*F*6) > ()
D Distinguish with CMI	*F*4,*F*6,*F*10,*F*16,*F*17,*F*20,*F*27	(*F*4,*F*6,*F*7,*F*12) > () (*F*16) > (*F*6)
P Distinguish with Boeing	*F*1,*F*4,*F*6,*F*10,*F*12,*F*14,*F*21	(*F*10,*F*16,*F*17,*F*20,*F*27) > (*F*6) (*F*6,*F*21) > (*F*1)
D Distinguish with Sandlin	*F*1,*F*10,*F*16,*F*19,*F*27	(*F*4,*F*6,*F*12,*F*14,*F*21) > (*F*1) (*F*1,*F*16) > ()
P Distinguish with Mineral Deposits Two	*F*1,*F*4,*F*16,*F*18,*F*21,*F*25	(*F*1,*F*10,*F*16,*F*19,*F*27) > () (*F*21) > (*F*1,*F*16)
D Distinguish witho Arc	*F*10,*F*16,*F*20	(*F*4,*F*18,*F*21) > (*F*1,*F*16)(*F*16) > ()
P Distinguish with Bryce	*F*1,*F*4,*F*6,*F*18,*F*21	(*F*10,*F*16,*F*20) > () (*F*6,*F*21) > (*F*1)
D Distinguish with Yokana	*F*7,*F*10,*F*16,*F*27	(*F*4,*F*6,*F*18,*F*21) > (*F*1) (*F*16) > ()
P Distinguish with Laser	*F*1,*F*6,*F*10,*F*12,*F*21	(*F*10,*F*16,*F*27) > () (*F*6,*F*21) > (*F*1)
D Distinguish with Robinson	*F*1,*F*10,*F*18,*F*19,*F*26	(*F*6,*F*12,*F*21) > (*F*1) (*F*1) > ()
P Distinguish with Den-Tal-EZ	*F*1,*F*4,*F*6,*F*21,*F*26	(*F*1,*F*10,*F*19) > () (*F*6,*F*21) > (*F*1)
D Distinguish with National Rejectors	*F*7,*F*10,*F*15,*F*16, *F*18,*F*19,*F*27	(*F*4,*F*6,*F*21,*F*26) > (*F*1) (*F*16) > (*F*15)

Neither Emery or FMC appear in the A* dialogue. In fact they are not particularly helpful, here it seems that the adversaries are feeling one another out and being as non-committal as possible.

The Defendant then *Distinguishes* FMC and *Analogises* with the CMI case. The two players now continue to *Distinguish* the previous case and *Analogise* another case until the 17th move in the theory when the Defendant starts making *Distinguish with Arbitrary Preference* moves. The Plaintiff responds by *Analogising* with the College Watercolor case which the Defendant the *Distinguishes* and states an *Arbitrary Preference*.

Table XXV. Explanation of adversarial Mason dialogue – part two

Theory move	Factors in cited case	Rules produced
P Distinguish with Valco-Cincinnati	*F*1,*F*6,*F*10,*F*12,*F*15,*F*21	(*F*10,*F*16,*F*19,*F*27) > (*F*15)
		(*F*6,*F*15,*F*21) > (*F*1)
D Distinguish with Arb Pref		(*F*6,*F*12,*F*15,*F*21) > (*F*1)
		(*F*1,*F*16) > (*F*6,*F*15,*F*21)
P Cite College Watercolor	*F*1,*F*15,*F*26	(*F*15) > (*F*1)
D Distinguish with Arb Pref		(*F*15,*F*26) > (*F*1)
		(*F*1,*F*16) > (*F*6,*F*15,*F*21)
P Cite SpaceAero	*F*1,*F*8,*F*15,*F*18,*F*19	(*F*15) > (*F*1)
D Distinguish with Arb Pref		(*F*8,*F*15,*F*18) > (*F*1)
		(*F*1,*F*16) > (*F*6,*F*15,*F*21)
P Cite Digital Development	*F*1,*F*6,*F*8,*F*15,*F*18,*F*21	(*F*6,*F*15,*F*21) > (*F*1)
D Distinguish with Arb Pref		(*F*6,*F*8,*F*15,*F*18,*F*21) > (*F*1)
		(*F*1,*F*16) > (*F*6,*F*15,*F*21)
P Cite KG	*F*6,*F*14,*F*15,*F*16,*F*18,*F*21,*F*25	(*F*6,*F*15,*F*21) > (*F*16)
D Distinguish with Arb Pref		(*F*6,*F*14,*F*15,*F*18,*F*21) > (*F*16)
		(*F*1,*F*16) > (*F*6,*F*15,*F*21)
P Cite Lewis	*F*1,*F*8,*F*21	(*F*21) > (*F*1)
D Distinguish with Arb Pref		(*F*18,*F*21) > (*F*1)
		(*F*1,*F*16) > (*F*6,*F*15,*F*21)
P Cite National Instrument	*F*1,*F*18,*F*21	(*F*21) > (*F*1)
D Distinguish with Arb Pref		(*F*18,*F*21) > (*F*1)
		(*F*1,*F*16) > (*F*6,*F*15,*F*21)
P Cite Reinforced	*F*1,*F*4,*F*6,*F*8,*F*15,*F*21	(*F*6,*F*15,*F*21) > (*F*1)
D Distinguish with Arb Pref		(*F*4,*F*6,*F*8,*F*15,*F*21) > (*F*1)
		(*F*1,*F*16) > (*F*6,*F*15,*F*21)

This continues to the end of the dialogue with the Plaintiff *Analogising* cases and the Defendant *Distinguishing* them and stating an *Arbitrary Preference* until the last Defendant move where it uses the *Problem Distinguish* move. The *Arbitrary Preference* that the Defendant states is always the same,($\{(F1, F16) \rightarrow D\} > \{(F6, F15, F21) \rightarrow P\}$), because all of the factors from Mason have been included in the theory and so the Defendant just keeps using the preference of the Defendant factors preferred over the Plaintiff factors. Essentially the Defendant is simply stating that he should win, and challenging the Plaintiff to show that this contention is untenable. This is not a bad strategy for a lost cause, since it can succeed unless the opponent has a sufficient stock of cases. This forces the plaintiff to produce more cases and to continually refine the theory. Even the last Defendant

Table XXVI. Explanation of adversarial Mason dialogue – part three

Theory move	Factors in cited case	Rules produced
P Cite Technicon	*F*6,*F*10,*F*12,*F*14,*F*16,*F*21,*F*25	(*F*6,*F*21) > (*F*16)
D Distinguish with Arb Pref		(*F*6,*F*12,*F*14,*F*21) > (*F*16)
		(*F*1,*F*16) > (*F*6,*F*15,*F*21)
P Cite Televation	*F*6,*F*10,*F*12,*F*15,*F*16,*F*18,*F*21	(*F*6,*F*15,*F*21) > (*F*16)
D Distinguish with Arb Pref		(*F*6,*F*12,*F*15,*F*18,*F*21) > (*F*16)
		(*F*1,*F*16) > (*F*6,*F*15,*F*21)
P Cite Trandes	*F*1,*F*4,*F*6,*F*10,*F*12	(*F*6) > (*F*1)
D Distinguish with Arb Pref		(*F*4,*F*6,*F*12) > (*F*1)
		(*F*1,*F*16) > (*F*6,*F*15,*F*21)
P Cite Goldberg	*F*1,*F*10,*F*21,*F*27	(*F*21) > (*F*1)
D Problem distinguish		(*F*1,*F*16) > (*F*6,*F*15,*F*21)

move with *Problem Distinguish* resorts to this rule preference. The end result of this dialogue is to prefer *F21-Knew Info Confidential* to *F1-Disclosure in Negotiation*. *F16-Info Reverse Engineerable* has already been established as losing to *F21-Knew Info Confidential* when in the presence of *F6-Security Measures* and *F15-Unique Product* after Televation has been cited.

For this dialogue the Plaintiff wins so the Defendant cannot refine the theory so as to produce a good theory for his side. The Defendant seems to run out of moves and just resorts to stubbornly using the *Distinguish with Arbitrary Preference* move in response to all of the Plaintiff's move, until it is able to insist on the preference only by appealing to a value preference, which terminates the dialogue.

For the Bryce dialogue the Plaintiff theory gets better results and so wins. The Plaintiff theory has better results than the A* dialogue in Bryce but the Defendant theory is worse than the A* dialogue in Bryce.

For the Ferranti dialogue the Defendant theory gets better results and so wins. In both cases it the outcome is so clear that a sophisticated theory is not required to silence the objections to the decision. From this we conclude that if the quality of the theory is important, it is essential to use a balanced seed case to give both sides the opportunity to develop a reasonable theory

13.3. ADVERSARIAL DIALOGUES IN ROSALIND

In ROSALIND the Adversarial search heuristic is still used but the two players can have different sets of background cases. This means that we can explore the effect of having different information available to the two players.

For dialogues taken from ROSALIND, we bias things in favour of the Plaintiff by giving the Plaintiff precedents containing a large number of factors and the Defendant precedents containing a small number of factors. These were labelled P1 in Section 12.

When using Adversarial search with these different case bases, we get the dialogues shown in Table XXVII. The theories obtained for the Plaintiff and Defendant players are identical except for the last move which is shown in bold type.

Because the players are restricted to a small set of cases that they can use, the dialogues are usually very small.

The Mason dialogues are longer than the other dialogues but only by one move, because Mason is a well balanced case and so the players have more moves that they can make. The dialogue for Mason is shown in Table - XXVIII. For this dialogue the Defendant player starts by *Analogising* with the Arco case, relying on *F16-Info Reverse Engineerable*, its best factor. This gives a rule of $\{(F16) \rightarrow D\}$ which is used to decide the Mason case but because there are no Plaintiff factors in Arco a rule preference is not produced. The Plaintiff responds by *Countering* with Technicon which gives a rule preference of $(\{(F6, F15, F21) \rightarrow P\} > \{(F16) \rightarrow D\})$. Since he has no effective cases available in his selection, the Defendant can only respond by *Distinguishing* Technicon and by stating an *Arbitrary Preference* which is $(\{(F16) \rightarrow D\} > \{(F6, F15, F21) \rightarrow P\})$ and is the reverse of the rule preference which the Plaintiff used from Technicon. This arbitrary preference results in a less good theory and so the Defendant loses and the Plaintiff wins.

For Bryce the Plaintiff wins by only using one move and the Defendant has to respond using the *Distinguish with Arbitrary Preference* which again gives a less good theory. For Ferranti the Defendant wins with one move so

Table XXVII. ROSALIND Search dialogues

Mason	Bryce	Ferranti
D Cite Arco	P Cite Boeing	D Cite Sheets
P Counter with Technicon	**D Distinguish with Arb Pref**	**P Problem distinguish**
D Distinguish with Arb Pref		

Table XXVIII. Explanation of ROSALIND Mason dialogue

Theory move	Factors in cited case	Rules produced
D Cite Arco	$F10$, $F16$, $F20$	$(F16) > ()$
P Counter with Technicon	$F7$, $F10$, $F15$, $F16$, $F18$, $F19$, $F27$	$(F6, F15, F21) > (F16)$
D Distinguish with Arb Pref		$(F6, F12, F15, F18, F21) > (F16)$ $(F16) > (F6, F15, F21)$

the Plaintiff has to respond with the *Problem Distinguish* move which also results in a less good theory.

When the cases are limited in this way there are not enough moves available to refine the theories to produce a theory which able to explain the background cases in a satisfactory way.

Conclusions

Heuristic search is necessary if we are to make full use of available background cases and the ability to use a more extensive background does improves the results for AGATHA. Moreover AGATHA produces better theories that the hand constructed theories reported in Chorley and Bench-Capon (2003a, b, c), and theories comparable in explanatory power to the best performing reported technique, IBP (Brüninghaus and Ashley 2003; Ashley and Brüninghaus 2003). Note also that AGATHA can be used even when there is no accepted structural model of the domain, whereas IBP relies on using the structure provided by the Restatement of Torts.

From our results we conclude that using all the cases is preferable to using only the most on point cases and that a depth factor 10% gives good results. Since our theories are not perfect, it might be possible to improve the evaluation used during the search by tuning the parameters. None the less we regard the performance as sufficient to indicate that the parameters and criteria used are at least in the right area.

We have also discovered that the best seed case to use for both the cooperative and adversarial modes of operation is one with a large number of factors and where these factors are divided equally between Plaintiff and Defendant. This case can, of course, be identified automatically, and so AGATHA can be used to construct a theory from a given background without manual guidance. It is important that a balanced case be used as the seed if a general explanatory theory is required. A case which is clear for one side of the other can be explained using a relatively simple theory, which does not address some of the more subtle interactions of factors required to give a theory which explains the domain in general.

We find the results reported here highly encouraging: they provide some support for the theoretical account of reasoning with cases in terms of theories which use factors and values proposed in Bench-Capon and Sartor (2003). Moreover they suggest that the process of theory construction may be open to automation, once the domain analysis required to produce the background has been carried out

The adversaries need as good a stock of cases as possible. While performance is not much affected if one side is unaware of the cases favouring the other side, they need to be able to make their own arguments to force their opponent to refine the theory. Where information is unbalanced the outcome of the dialogue is not much affected by which side has the better information: often a better theory is produced if the side with the better case has the worse information.

Given these conclusions, adversarial search produces theories comparable in performance to A*, but is more efficient in terms of nodes examined during the search.

A second aim of our work was to construct these theories through the use of dialogues. Our conclusions regarding this are:

When the non-adversarial heuristic search is used, we get a sequence of cases which can be explained in terms of plausible domain arguments.

When adversarial search is used, there is a tendency to use arbitrary preferences, hypothesising theories that are not grounded in any cases. This reflects the desire of the adversary to avoid strong moves from their opponent.

In practice this delays rather than prevents the use of significant cases.

This delay is more than compensated for by the narrower tree generated in adversarial search.

Where the case is clear for one side or the other, the search will terminate with a theory which meets the current case, but which does not generalise.

In summary we have presented a series of experiments which confirms the potential of viewing reasoning with legal cases as theory construction, and developed a set of tools which provide promise for automating this process.

Notes

[1] Goldberg, National Instrument, Ecologix and Sandlin.

[2] SpaceAero and National Rejectors.

[3] Trandes and CMI.

[4] Mineral Deposits, was modified to add a factor reflecting that the product was loaned in confidence by the plaintiff to the defendant. This seems important, e.g. "compare Mineral deposits Ltd. v. Zigan, 773 P. 2d 606 (Colo. App. 1988) (reverse engineering not allowed when product loaned in confidence)" (Lipinski and Britz 2000). We refer to this modified case as Mineral Deposits Two.

References

Aleven, V. (1997). Teaching Case Based Argumentation Through an Example and Models. PhD Thesis, The University of Pittsburgh.

Ashley, K. D. (1990). Modelling Legal Argument. Bradford Books, MIT Press: Cambridge, Mass.

Ashley, K. and Brüninghaus, S. (2003). A Predictive Role for Intermediate Legal Concepts: In Bourcier, D. (ed.), Proceedings of Jurix 2003. IOS PressAmsterdam, 153–162.

Bench-Capon, T. and Sartor, G. (2003). A Model of Legal Reasoning with Cases Incorporating Theories and Values, Artificial Intelligence 150(1–2): 97–143.

Branting, K. L. (1991). Reasoning with Portions of Precedents. In Proceedings of the Third International Conference on AI and Law, 45–54. ACM Press: New York.

Brüninghaus, S. and Ashley, K. D. (2003). Predicting Outcomes of Case-based Legal Arguments. In Proceedings of the Ninth International Conference on AI and Law, 233–242. ACM Press: New York.

Chorley, A. and Bench-Capon, T. (2003a). Developing Legal Knowledge Based Systems Through Theory Construction. Technical Report ULCS-03-013, Department of Computer Science, The University of Liverpool. http://www.csc.liv.ac.uk/research/techreports/tr2003/ulcs-03-013.pdf.

Chorley, A. and Bench-Capon, T. (2003b). Developing Legal Knowledge Based Systems Through Theory Construction. In Proceedings of the Ninth International Conference on AI and Law. ACM Press: New York.

Chorley, A. and Bench-Capon, T. (2003c). Reasoning with Legal Cases as Theory Construction: Some Experimental Results: In Bourcier, D. (ed), Proceedings of Jurix 2003. IOS PressAmsterdam, 173–182.

Chorley, A. and Bench-Capon, T. (2004a). Support for Constructing Theories in Case Law Domains: In Galindo, F., Takizawa, M. and Traunmuller, R. (ed), Proceedings of DEXA 2004. LNCS. 3180 Springer VerlagBerlin, 508–517.

Chorley, A. and Bench-Capon, T. (2004b). AGATHA: Automation of the Construction of Theories in Case Law Domains. In Proceedings of Jurix 2004, 89–98. IOS Press: Amsterdam.

Chorley A. and Bench-Capon T. (2005). AGATHA: Automated Construction of Case Law Theories Through Heuristic Search. In Proceedings of the Tenth International Conference on AI and Law. ACM Press: New York.

Chorley, A. and Bench-Capon, T. (to appear). An Empirical Investigation of Reasoning with Legal Cases Through Theory Construction and Application. Artificial Intelligence and Law.

Lipinski, T. A. and Britz, J. J. (2000). Rethinking the Ownership of Information in the 21st Century: Ethical Implications, Ethics and Information Technology 2, 49–71.

Prakken, H. and Sartor, G. (1998). Modelling Reasoning with Precedents in a Formal Dialogue Game, Artificial Intelligence and Law 6, 231–287.

Skalak, D. B. and Rissland, E. L. (1992). Arguments and Cases: An Inevitable Intertwining, Artificial Intelligence and Law 1, 3–44.

Winston, P. H. (1992). Artificial Intelligence. (3rd ed.). Addison Wesley: Reading, Mass.