

# Semantic Negotiation: Modelling Ambiguity in Dialogue

Alison Pease                  Simon Colton

Alan Smaill

University of Edinburgh, Division of Informatics,  
80 South Bridge, Edinburgh, EH1 1HN, Scotland

John Lee

University of Edinburgh, Division of Informatics,  
2 Buccleuch Place, Edinburgh, EH8 9LW, Scotland

email: {alisonp@dai, simonco@dai, smaill@dcs, john@cogsci}.ed.ac.uk

## Abstract

We argue that negotiation over the meaning of terms in a statement is part of human discussion and that it can lead to richer theories. We describe our preliminary model of semantic negotiation and discuss theoretical examples which we hope to implement. Finally we consider how semantic negotiation fits into existing work on argumentation.

## 1 Introduction

There many situations in which participants realise partway through an argument that they are interpreting one of the key terms differently. The meaning of the key term is then called into question, and the focus of the argument switches from the truth or acceptability of a claim or offer to the *meaning* of the term. This is particularly common in subjects such as philosophy, law and politics, in which persuasive reasoning is all important and concept definitions are modified according to the proponents' goals. Yet this phenomenon is rarely seen in AI research. Economic agents may well disagree on the price of a potato, even haggle over it in a reasonably sophisticated way – but they never start arguing about what a potato is.

In this paper, we argue that:

- 1) people sometimes define concepts in a way which supports their beliefs or goals (§2.1), and
- 2) subsequent disagreement over the meaning of a concept can result in a richer theory (§2.2).

We are currently modelling this phenomenon, (described in §4), with the aim of (a) elucidating it and (b) using it to extend existing theory forma-

tion programs. The fact that although participants in a discussion use a shared language, some of the terms are ambiguous, raises questions like – what sorts of things can be ambiguous? How might ambiguity arise? How can it be resolved? Can it be used to produce richer theories? By modelling *semantic negotiation* – the problem of reaching mutually acceptable definitions (a specialised type of *negotiation*, defined by the agent community as the problem of reaching mutually acceptable agreements) – we hope to address these questions.

We draw on work by Jennings et al. (1998) (described in §3), and in §5 place our ideas in the context of previous work on argumentation.

## 2 Semantic negotiation and its value

### 2.1 Ambiguity in human reasoning

The failure at the beginning of the last century of the quest for a perfect language in which neither ambiguities, nor paradoxes or redundancies exist, showed the difficulties involved in writing a formal language which can be used to describe a reasonably large domain. Today it is normally accepted that any non-trivial language is likely to contain ambiguities. Different types of ambiguity include *lexical* (where a word has two different meanings); *syntactic* (a sentence with two syntactically correct derivation trees which indicate different meanings); *semantic* (a sentence with two meanings, only one of which makes sense), and *pragmatic* (the meaning of a word is relative to the speaker). Much work on AI and am-

biguity is concerned with methods to automatically determine a writer's intended meaning (for example Romacker and Hahn (2001) who focus on representing and managing ambiguity in natural language text understanding). In contrast, we are concerned with ways in which ambiguity may be exploited (or introduced into a previously unambiguous concept), in order to support an argument or set of beliefs.

Examples in which ambiguity is used to support an argument include an insurance company which argues that a house damaged in a hurricane is not covered by the owner's accident policy, as a hurricane is an 'act of God' rather than an accident. Similarly, a lawyer may argue that a client who jumped a red light while rushing his wife to a maternity ward is not guilty of reckless driving. In the 1990's the European Union Food Standards discussed the definition of chocolate. The minimum cocoa content which a substance must contain in order to be called chocolate was debated, as countries which produce it with a higher cocoa content did not want their chocolate to be confused with products with a lower cocoa content. Finally, consider the recent controversy over the meaning of 'prisoners of war'. When challenged that their treatment of the Taliban prisoners violated the Geneva Convention – that all prisoners of war should be treated humanely – the American government argued that the prisoners were not 'prisoners of war', but 'battlefield detainees'<sup>1</sup>. In these examples, the terms *accident*, *reckless driving*, *chocolate* and *prisoners of war* are defined by each party in such a way as to aid their argument.

## 2.2 Using semantic negotiation to enrich a theory

We discuss three examples of semantic negotiation in mathematics, and show how it has enriched mathematical theories.

### The concept 'polyhedron'

Lakatos (1976) shows both *that* ambiguity exists even in mathematics, and – of key importance to AI researchers – he shows *how* it might arise and how it can be used effectively. He presents a dialogue between a group of students and their teacher, in which the history of Euler's conjecture and its proof is enacted as a rational reconstruction. As well as providing a rare insight into the way in which theories

<sup>1</sup>The definition of *humane treatment* was also disputed – in particular whether it could ever include interrogation, as the American government felt it important to interrogate the prisoners while not wanting to be open to the charge of inhumane treatment.

evolve (where by theory we mean concepts, counterexamples, conjectures and 'proofs' or arguments), this work is of great value to researchers modelling dialogue. Argumentation moves are set out which Lakatos categorises into various methods.

Euler's conjecture states that for all polyhedra, the number of vertices (V) minus the number of edges (E) plus the number of faces (F) equals 2. Starting from this conjecture, one of the students finds the counterexample of the hollow cube – a cube with a cube-shaped hole in it (figure 1).

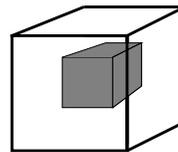


Figure 1: The hollow cube;  $V - E + F = 16 - 24 + 12 = 4$

One reaction to this counterexample is to question the meaning of the concept 'polyhedron'. Rather than accept it as a counterexample, it is branded a *monster* since, it is claimed, it is not a polyhedron. Participants in the discussion then argue about the definition of polyhedron. Lakatos (1976) calls this process of arguing that a counterexample is not valid and therefore not a threat to a conjecture, the method of *monster-barring*. Using this method repeatedly participants expand the theory of polyhedra, differentiating between 'a solid whose surface consists of polygonal faces' to 'a surface consisting of a system of polygons' (thus excluding the hollow cube), to 'a system of polygons arranged in such a way that (1) exactly two polygons meet at every edge and (2) it is possible to get from the inside of any polygon to the inside of any other polygon by a route which never crosses any edge at a vertex'. They also discuss the meaning of the concepts polygon, area and edge. As a result definitions are tightened and students gain a greater understanding of the conjecture.

### The concept 'number'

In the late 19th century Cantor introduced the mathematical community to the 'number'  $\aleph_0$ , which is the size of the set of all integers (the first transfinite number). This was a counterexample to the conjecture that if you add a non-zero number to another number, then the second number always changes, since  $\aleph_0 + \aleph_0 = \aleph_0$ . Similarly it violates the conjecture that any positive number multiplied by 2 is bigger than the number, as  $\aleph_0 \cdot 2 = \aleph_0$ . The

law of monotonicity, that for all numbers  $a$ ,  $b$  and  $c$ , if  $b < c$ , then  $a + b < a + c$  fails if  $a = \aleph_0$  (for any finite  $b$  and  $c$ ). For these and other reasons, initial reaction to Cantor's work was hostile,  $\aleph_0$  branded a monster (to use Lakatos' terms), and barred from the concept of number. However Cantor was developing a whole area of mathematics which he considered to be interesting and worthwhile – which included the number  $\aleph_0$ . Therefore he continued his research and tried to convince the mathematical community (eventually with success) that transfinite numbers *are* a kind of number and a valid area of mathematics. This process of suggesting that a new sort of object *is* a number, initial denial and later acceptance when it proves its worth can be seen repeatedly in number theory. For instance the number zero was initially branded a monster by the Greeks, for various reasons including its violation of the conjecture that if you add a number to another number, then the second number always changes. When zero was accepted this conjecture was modified to that above (*i.e.*, excluding zero). Other examples of ambiguity in the concept of number include initial barring of 1 (barred by the Pythagoreans as it challenged their belief that all numbers increase other numbers by multiplication);  $\sqrt{2}$  (it violated the Greek belief that all numbers describe a collection of objects); and  $x = \sqrt{-1}$  (violating the law of trichotomy, for any numbers  $x, y$ , either  $x = y$  or  $x < y$  or  $x > y$ ). Now of course 0, 1, irrational and imaginary numbers are accepted without question, and the concept of number has been generalised to complex numbers and beyond (quaternions). Clearly number theory (and other areas of mathematics) have been greatly enhanced by these additions.

### The concept 'prime'

Another example in number theory is the definition of prime number. A prime was initially defined to be 'a natural number which is only divisible by itself and 1'. However this definition includes the number 1, which was found to be a counterexample to many theorems and conjectures about primes. For instance, the Fundamental Theorem of Arithmetic (FTA) states that every natural number is either prime or can be expressed uniquely as a product of primes. If 1 is also considered prime then this violates the uniqueness claim, since, for example,  $6 = 2 * 3 = 2 * 3 * 1 = 2 * 3 * 1 * 1 = 2 * 3 * 1 * 1 * 1 = \dots$  Rather than explicitly exclude the prime 1 from this (and other) theorems, it is preferable to exclude it from the concept definition, and today the accepted definition of a prime is 'a natural number with ex-

actly two divisors'. This change in the concept definition has enabled many theorems about primes, including the FTA, to be neatly stated.

## 2.3 Properties of semantic negotiation

By extracting some general properties from the examples above, we can begin to answer the questions we asked in §1, and to see how we might model this phenomenon.

### What sorts of things can be ambiguous?

Some ambiguous entities are (sub)concepts within a universe of objects (such as a prime), and some are the main concept, *i.e.*, the universe itself (such as a polyhedron, or number).

### How might ambiguity arise?

Some concepts are initially specifically defined (everyone agrees on the definition), and the definition changed (someone argues that a second definition would be more useful). Others are initially vague (it is not known whether some objects are examples of the concept or not) and only when disagreement arises are different concept definitions made explicit.

### How can ambiguity be resolved?

Experts evaluate the worth of each of the rival definitions (often with different results). The existing definition is assumed by default, with the onus on proponents of a new definition to convince the other experts of its value. Grounds for accepting the new definition include showing that it produces interesting new theories or results (including preserving the 'truth' of a faulty conjecture). There is usually a period during which it is unclear whether the object in question belongs to a concept or not. It passes through a period of indefinite status with some people accepting its status, others not, others unsure, until it either proves its worth and is generally accepted or fails to convince enough people and gradually disappears.

Lakatos (1976) does not explore reasons for choosing one concept definition over another. Instead the teacher in the dialogue simply asks everyone to accept the strictest, *i.e.*, most limited definition suggested so far (at least for the duration of the discussion). However the only clear end to this process is a tautology (hence a student's sarcastic suggestion that a polyhedron be defined as 'a system of polygons for which the equation  $V - E + F = 2$  holds' – (Lakatos, 1976, p. 16). Dunmore (1992) suggests that concept definitions are chosen and developed

according to their use. For instance a concept which is not associated with any interesting conjectures is unlikely to become well known and accepted within the mathematical community.

### 3 Argumentation-based negotiation

Jennings et al. (1998) emphasise the importance of negotiation in multi-agent research, and outline an informal framework describing its key features. They divide negotiation issues into *protocols* (rules which govern interaction), *objects* (the range of issues over which agreement must be reached) and *Agents' Decision Making Model* (the way in which an agent follows the protocol to achieve its objectives).

Agents move through the space of possible agreements (in our case all the candidate definitions for a concept), defining their own spaces of acceptable points. Negotiation works by agents suggesting points in the space which are acceptable, and evaluating each point suggested. This ranking may change during negotiation, as agents are persuaded that a point is valid. The way in which they rate points may also change. A minimal requirement is that agents be able to propose some part of the agreement space as acceptable, and can respond to other agents' suggestions. A more efficient model would give agents capability to explain *why* they are rejecting/proposing a certain point. This might include rejecting a proposal but stating which aspects were considered good, a *critique*, or making a *counter-proposal* in response to a proposal. Such a model might include *justifications* – in which the agent states its reasons for making a proposal, or *persuasion* – in which an agent tries to change another's agreement space or rating over the space. These arguments help to support an agent's stance.

Jennings et al. (1998) state that an agent capable of argumentation-based negotiation must have a mechanism for:

- communicating proposals and supporting arguments;
- generating proposals;
- assessing proposals and arguments; and
- responding to proposals.

They do not suggest that the meaning of a con-

cept could be a possible object over which agents negotiate, giving as examples issues relating to negotiation over services or products. However, the framework is general enough to include this type of negotiation. For instance, once the disagreement over a concept definition has arisen, an agent might suggest one and justify why it is good (for example it may be used in many of its conjectures) - which may lead another agent to re-evaluate the definition and rate it more highly.

### 4 A preliminary model of semantic negotiation

We are currently extending the HR system (Colton, 2001) to perform semantic negotiation. In this section we briefly outline the original system, as well as the extended version. We then describe two theoretical examples of semantic negotiation and ideas on their implementation in HR.

#### 4.1 The HR system

HR is an automated theory formation program (Colton, 2001) which is given background information about a domain, including some objects of interest (the 'universe') – such as integers – and some initial concepts – such as multiplication and addition. It forms new concepts by using one of 10 general production rules to transform one (or two) old concepts into a new one. The examples of a concept are used to make conjectures empirically.

HR is able to evaluate concepts and conjectures based on various interestingness criteria, which measure values such as the *novelty* of a concept, the *number of open or true conjectures which the concept is in*, and the *surprisingness* of a conjecture.

#### 4.2 Extensions to HR

In order to model semantic negotiation (and as part of a project to model Lakatos-style reasoning) we have partially implemented an agent architecture in which multiple copies of HR are run which can communicate details of their theory to each other. These consist of 'students' and 'teacher', and all have different weightings of interestingness values, for example one may favour novel concepts while another prefers surprising.

Lakatos (1976) identifies various reactions to a counterexample to a conjecture, and in §2.2 we described one reaction – the method of monster-barring. Another reaction may be to modify the

conjecture by barring the counterexample. For instance, we may generalise from the hollow cube to ‘polyhedra with cavities’, and then modify Euler’s conjecture to *for all polyhedra without cavities,  $V - E + F = 2$* . This is the method of exception-barring, which we have implemented in the extended version of HR, and expect to use in our model of semantic negotiation.

Using the reflection mechanism available in Java – which enables information about classes and individual objects to be obtained at run-time – we have implemented a cut-down Java interpreter in HR, which can be used to interpret and execute Java code at run-time. Currently, the interpreter can handle the creation of new objects and various constructs, including for loops, if-then-else statements and string manipulations. We intend to extend the functionality of the interpreter to handle more complicated Java code.

Having a run-time interpreter has various advantages, including:

- homogeneity in the code used to control HR at compile time and at run-time;
- easily compiling pieces of run-time code to become compiled parts of HR, which improves efficiency;
- introspective access to the main ‘theory’ object which HR has built, which enables it to run any methods attached to that object, and to build and execute new functions at run-time.

Various aspects of the way HR forms and presents theories take advantage of the interpreter, including the way it reports theories to the user, and the reaction mechanism, which is a particular heuristic search whereby HR reacts to certain events in the theory formation by taking various additional theory formation steps. The reaction scripts are written in Java and describe which events to react to, and how to react.

In terms of the project discussed here, HR’s Java interpreting functionality can be used to avoid an inter-lingua in the communication between agents. That is, the ways in which the various agents instruct, inform and control each other do not have to be prescribed before each run is undertaken. This will enable a more flexible communication between the agents.

### 4.3 Two theoretical examples

Suppose that *alpha*, *beta* and *gamma* are different versions of HR which have been running indepen-

dently for a time. We are aiming to generate the sort of dialogue shown here. Note that, as described above, they would communicate in java code rather than natural language – here we paraphrase to aid the example. Superscript numbers indicate places where we discuss ideas for implementing aspects of the discussion.

#### Example one: proposing to exclude an existing object

*alpha*: I’ve noticed that the object *o* is a counterexample to many of my conjectures about concept *C* (*C* may be the universe or a concept)<sup>1</sup>. It would be neater to bar it from the concept rather than my conjectures. What do you all think?<sup>2</sup>. Does anyone have a neat definition which includes exactly those objects in *C* except for *o*?

*beta*: It’s a counterexample to many of mine as well. Let’s bar it.

*gamma*: I disagree. It does not violate any of my conjectures. For which conjectures of yours is it a problem?

*alpha*: For conjectures *X*, *Y* and *Z*.

*beta*: Also conjectures *U*, *V* and *W*.

*gamma*: OK these are interesting conjectures. Let’s revise our concept definition then<sup>3</sup>. I have a concept *C*<sub>1</sub> which includes the same objects as *C* except *o*. Let’s use this from now on<sup>4</sup>

#### Implementing the dialogue

(1) If an agent has many conjectures of the type  $\forall x \neq n, P(x) \rightarrow Q(x)$  in its theory then it may propose a concept definition change. Whether the object in question is to be excluded from the universe or from a concept depends on whether the problem conjectures all involve the *same* concept (in which case exclude the object from this concept), or different concepts (in which case exclude it from the universe).

(2) We now have an ambiguity since it is unclear whether a concept includes a given object or not. In order to decide, we need (i) a way for individual agents to determine their position, and (ii) a negotiation protocol to determine how agreement is reached. We discuss these below.

(3) any agent can look in its theory to see if it has a concept which covers certain objects.

(4) HR has a mechanism – replace definition – which is currently used to overwrite old definitions if a simpler one is found. The more complex definition is still in the theory (and in all the conjectures prior to the new definition), but is no longer referred to. We can use this mechanism to replace the old definition

with a new one.

(i) Deciding individually whether to exclude an object

Currently the interestingness measures in HR evaluate concepts and conjectures. We can extend these to evaluate how interesting an object is. For instance, we may measure:

- generality of conjectures – the number of conjectures in the theory which involve the concept in question, which the object does *not* violate (the more general a conjecture is, the more valuable);
- counterexample-barred conjectures – the inverse of the number of counterexample-barred conjectures in the theory which involve the concept in question to which the object is a counterexample.
- broken conjectures – the number of conjectures in the theory which are violated by the object.

Of two rival definitions, an agent might prefer the simplest. Colton has already implemented this in HR. It arises when a conjecture that two concepts are equivalent has been proven. HR then evaluates them both to determine the simplest, where simple is defined as the number of production rules used to generate a concept (the fewer the simpler). The user can instruct HR to always keep the simplest definition, with the more complicated definition only appearing in equivalent conjecture statements. An agent might also judge which of two rival concepts is more interesting according to its interestingness criteria already present.

Using these (and other) measures, an agent can evaluate and respond to a proposal to exclude an object – which it may or may not already have in its theory. Clearly, these values will be different for each agent, as they will have a different weighting of interestingness values, so would evaluate the same object or concept differently. Additionally they have different theories – *i.e.*, different conjectures and concepts against which the object or concept is measured. As with other interestingness measures in HR, we anticipate making these flexible (a weighted combination input by the user) and then experimenting to see which combination is the most productive.

(ii) A negotiation protocol

Following Jennings et al. (1998), the space of all possible agreements is all candidate definitions for a concept. Agents define their own space of acceptable points by evaluating their own candidate definitions (those which score the highest according to their interestingness values). Negotiation begins by

the agents suggesting points in the space which are acceptable, and evaluating each received definition (which may score more highly than one which they have generated themselves). This constitutes the simple model described in (Jennings et al., 1998). We would then want to enable agents to communicate the reasons they reject or propose a definition, for instance “I reject definition D because it excludes the number 1 and I need this number for all my proofs”, or “I propose definition D because it would preserve conjecture C, which is an interesting one”. Definitions could then be re-evaluated, for example if the agent receiving the latter proposal also evaluates conjecture C as interesting, it might be persuaded that definition D is more interesting than the one it previously proposed.

The agents re-evaluate their measures when others present reasons for/against accepting a concept definition.

This dialogue demonstrates how we may automate ambiguity in both the universe of objects and (sub)concepts, a period of indefinite status while it is discussed and the sorts of arguments which might be made for changing to a new definition or sticking to an existing one, and methods by which everyone might agree on a new definition and then change to it.

A similar dialogue can be envisaged in which a proposal to accept an object, *i.e.*, to *widen* rather than narrow a concept definition to include rather than exclude it, is made. If agreement is not reached then the teacher could decide, based on a weighting of the results of its own interestingness criteria and those of the students’. It may also put more weight on the existing definition (since it is costly to revise definitions). The students would then use the specified definition, at least for the duration of the discussion.

### Example two: object-driven concept formation

Suppose that a conjecture has been proposed, which is true for all numbers between 1 and 100 except 1, 17 and 72. We may wish to find a concept to cover these counterexamples (and possibly others which lie outside the range), *i.e.*, a ‘concept-to-exclude’. Therefore the teacher may ask the students to find such a concept. Each would then look in its theory and send a specific (possibly different) concept back. The ‘concept-to-exclude’ is now ambiguous, and agents rate the definitions and negotiate over which is the best.

This example encompasses:

- 1) an initially vague concept – it is not known

- whether some objects are examples or not;
- 2) the vague concept developed independently into (possibly) different explicit concepts;
  - 3) each agent suggesting arguments for accepting its concept.

## 5 The ‘fallacy’ of ambiguity — where does it fit into argument analysis?

In order to place our work in context, we conclude by briefly describing approaches to argumentation and ambiguity, and noting our position towards these perspectives.

Aristotle (1957, 1955, 1976) identified three motivations behind arguments; *apodictic*, *dialectic* and *rhetoric*, in which certainty, a general acceptance and convincing an audience is respectively sought. Although many would claim that motivation behind mathematical argument falls into the first category, the second (or even third) is more appropriate to mathematics as Lakatos (1976) describes it. Aristotle treated dialectical argument as a game between a defender and attacker, and suggested guidelines such as forcing the defender to contradict herself, state an untruth or paradox, or defend a circular argument, for conducting the debate. Certain moves – called fallacies – were disallowed, including the fallacy of ambiguity.

In his work on controversy, Crawshay-Williams (1957) emphasised the need for clarification of concepts prior to discussion. He claimed that if participants in a discussion agree upon the criteria under which a statement will be tested, then agreement regarding its absolute/probable/indeterminate truth will soon be reached. He calls one such criterion *conventional*, to mean the condition that participants agree on the meaning of terms (the others are *logical*, in which inference rules must be agreed, and *empirical*, in which facts and their contextual description should also be agreed).

Naess (1953) also stated that criteria for the verification or falsification of a statement are essential (to the extent that if no such criteria are found then discussion should be abandoned). However he includes agreeing on terms as a stage in the discussion, rather than a pre-requisite to it. The three stages in resolving a discussion, he suggests, are interpretation, clarification and argumentation. For any statement  $T$  there is set of possible interpretations of  $T$ , and participants must agree on which interpretation they wish to discuss. He claims that

*precizing* statements, (being more precise) helps to eliminate misunderstandings, where  $U$  is more precise than  $T$  if any interpretations of  $U$  are also interpretations of  $T$ , but there are interpretations of  $T$  which are not interpretations of  $U$ . This is useful only if the disagreement has occurred through different interpretations, he does not advocate continual precization of statements since discussion would be practically impossible. Disagreements rooted in misunderstandings are termed *verbal*, only if after there is still disagreement after precization is it *real*. In the case of a real disagreement the evidence is weighed up to see which of the two statements is more acceptable.

Carbogim et al. (2000) present a survey of issues which can be handled by automated argumentation systems and suggest directions for future research. They consider the generation and evaluation of arguments, including issues such as drawing conclusions from an incomplete or inconsistent knowledge base, decision making under uncertainty and multi-agent negotiation systems. Argument about the meaning of terms used is not considered either explicitly in the text, nor in any of the examples, although it may turn out that semantic negotiation fits into one of the frameworks outlined.

Our view is that ambiguity plays an important role *within* a discussion. That is, questioning the meaning of terms in a discussion *is* a valid strategy (with restrictions on which words may be questioned). We differ from Crawshay-Williams’ approach in that we see debate of terms as an important part of discussion rather than a pre-requisite of it. Participants may not realise initially that they have different interpretations of a word, indeed they may not themselves have a clear interpretation. We also differ from Naess’ linear approach, disagreement over terms could arise at any point in a discussion. Of particular interest in Naess’ work is the *evidence* used to resolve ‘real’ disagreements, and we hope by implementing semantic negotiation to elucidate the sort of evidence required. While the survey by Carbogim et al. (2000) is not intended to be exhaustive, it does cover the central issues. The omission of semantic negotiation suggests that it is either irrelevant to argumentation or is a new direction which has received little attention. We hold that it is the latter.

## 6 Conclusion

We have argued that negotiation over the meaning of terms in a statement is part of human argument

and that it can lead to richer theories. We have also described our preliminary model of semantic negotiation and discussed theoretical examples which we hope to implement. Finally we have considered how it fits into existing work on argumentation. We consider this phenomenon to be a fertile and relevant area of research in the fields of argumentation and agent-based negotiation.

Our goal now is to extend the HR system (Colton, 2001) to model semantic negotiation. Issues which we expect to address include:

- how one definition is chosen over another. In Lakatos (1976) the narrowest definition is chosen. Other criteria may be the most common definition, or the most interesting;
- how are different interpretations best represented?
- what appropriate argumentation forms are there for this form of negotiation?
- to what extent is it useful to question word meaning?

We also intend to apply the Lakatos-style reasoning enabled by the agency to machine learning problems. In particular, we hope the agency could take over where machine learning programs currently finish. For example, suppose a program has learned a way of predicting whether chemicals are carcinogenic, based on their molecular structure. If the prediction has a success rate of, say, 85%, then 15% of the chemicals break the prediction rule. If we collect these together as the examples of a new concept, which we define as: “The set of chemicals which break the prediction rule”, then we have a suitably ambiguous concept to which the agency can apply their negotiation techniques. If the agency could identify a definition for this concept, then it is possible that the additional information could be incorporated into the prediction rule itself to increase accuracy, or at least provide a better understanding of the domain.

## Acknowledgements

We would like to thank the two anonymous reviewers for some very helpful comments on a previous draft. This work was supported by EPSRC grants GR/M45030 and GR/M98012. The second author is also affiliated with the Department of Computer

Science, University of York.

## References

- Aristotle (1955). *On sophistical refutations*. Harvard University Press, USA.
- Aristotle (1957). *Aristotle’s prior and posterior analytics: a revised text*. Clarendon Press, Oxford.
- Aristotle (1976). *Rhetoric*. Berolini.
- Carbogim, D., Robertson, D., and Lee, J. (2000). Argument-based applications to knowledge engineering. *The Knowledge Engineering Review*.
- Colton, S. (2001). *Automated Theory Formation in Pure Mathematics*. PhD thesis, Dept. of Artificial Intelligence, University of Edinburgh.
- Crawshay-Williams, R. (1957). *Methods of Criteria of reasoning: An Inquiry into the Structure of Controversy*. London: Routledge and Kegan Paul.
- Dunmore, C. (1992). Meta-level revolutions in mathematics. In Gillies, D., editor, *Revolutions in Mathematics*, pages 209–225. Clarendon Press, Oxford.
- Jennings, N. R., Parsons, S., Noriega, P., and Sierra, C. (1998). On argumentation-based negotiation. In *Proceedings of the International Workshop on Multi-Agent Systems*, Boston, USA.
- Lakatos, I. (1976). *Proofs and Refutations*. CUP, Cambridge, UK.
- Naess, A. (1953). *Interpretation and Preciseness: A Contribution to the theory of Communication*. Oslo: Skrifter utgitt ar der norske videnskaps academie.
- Romacker, M. and Hahn, U. (2001). Context-based ambiguity management for natural language processing. In *Proceedings of the Third International Conference on Modelling and Using Context*, pages 184 – 197, Dundee, Scotland.