A Conversational Agent System as a Test-Bed to Study the Philosophical Model "DC"

Tangming Yuan and **David Moore** and **Alec Grierson**

School of Computing,
Leeds Metropolitan University
Leeds LS6 3QS
United Kingdom

Abstract

This paper reports research concerning a suitable dialogue model for human-computer debate on a controversial issue such as capital punishment. We consider the adoption of Moore's [1993] utilization of Mackenzie's [1979] game DC, and in particular means of building conversational agents as the test-bed to facilitate evaluate of certain aspects of the proposed model. This study reveals several weaknesses of DC in preventing fallacious and common errors. It is anticipated that this work will contribute toward the development of human computer dialogue, and help to illuminate research issues in the field of dialectics itself.

1 Introduction

A previous paper [Yuan et al., 2002] considered the use of a computational dialectical approach as a means of providing a suitable model for an educational human-computer debating system. Here, we seek to further that investigation in several ways. First, we rehearse the argument for the adoption of Moore's [1993] utilization of Mackenzie's [1979] game DC, and provide a brief introduction to the model. We then discuss means of evaluating the proposed model in preventing fallacies and common errors, and construct two conversational agents capable of engaging in debate with each other via the proposed model. These agents, we argue provide a test-bed for facilitating evaluation of certain aspects of a dialectical system. Finally, we analyze the agent-generated dialogue and categorize problems concerning the DC rule set.

2 Moore's Utilization of DC

Following Moore [1993] and Walton [1989], it can be argued that a set of criteria is required for a suitable underlying model for a human-computer debate:

- the game needs to be persuasion style and symmetric in nature.
- the set of move types provided should be adequate for expression.
- the protocol should leave enough room for strategic formation.
- the protocol should be able to prevent fallacious argument.
- the model should be computational tractable.
- it should raise light cognitive load to the user.

Using these criteria, we have conducted a comparative study of the most recent development of dialogue models (e.g. [Prakken, 2000]; [Bench-Capon, 1998]; [Walton and Krabbe, 1995]; [Ravenscroft and Pilkington, 2000]; [Moore, 1993]; [Lodder and Herczog, 1995]) in the area of informal logic and computational dialectics. The study suggests that Bench-Capon's system is explanation based. Prakken's framework and Lodder and Herczog's systems are based on non-monotonic reasoning, and may be suitable for domains like legal and scientific proof where the strength of arguments can be predefined or arbitrated by a judge. However, in a controversial domain such as capital punishment, the strength of argument may be based largely on participants' judgement and is difficult to pre-specify. Further, their systems lack a question move type, and this may prevent students from asking the tutor questions and tutors from questioning the student's understandings, and this is undesirable from educational point of view (cf. Veerman [2002]). Ravenscroft and Pilkington's system is asymmetrical, and the dark side commitment of Walton and Krabbe's PPD would raise the cognitive load to the user. This is not of course to deny the general worth of these systems, but rather to suggest that they may not be a perfect match of the specific requirements of educational human computer debate. However, DC arguably meets most of the requirements. Further advantages of adopting DC as the underlying dialogue model are discussed by Moore and Hobbs (1996): the design of rules increases the computational tractability, and its symmetric nature enables either the tutor or the students to build their own positions. Further, Walton (1984) also suggests that the set of DC rules is practically useful. For the reasons discussed above, Mackenzie's (1979) game DC is chosen as the base system for further study.

The amended version of DC is specified in [Moore and Hobbs, 1996] as follows. There are five move- types:

(i) Statements. P, Q, etc. and the truth-functional compounds of statements: 'Not P', 'If P then Q', 'P and Q'; (ii) Questions. The question of the statement P is 'Is it the case that P?'; (iii) Challenges. The challenge of the statement P is 'Why is it supposed that P?'; (iv)Withdrawals. The withdrawal of the statement P is 'No commitment P'; (v) Resolution demands. The resolution demand of the statement P is 'resolve whether P'. There are five commitment rules: (i) CR_o: the initial commitment of each participant is null; (ii) CR_w: after the withdrawal of P, the statement P is not included in the speaker's store; (iii) CR_s: A statement P results in P being added to each store; (iv) CR_{ys}: After a statement P, if the preceding event was 'Why Q?', P and 'If P then Q' are included in each store; (v) CRy: A challenge of P results in P being added to the store of the hearer, and P being removed from, and 'Why-P?' being added to, the store of the maker of the move. Six dialogue rules are specified: (i) R_{form}: Participants may utter individual permitted locutions in turn; (ii) $R_{repstat}$: Mutual commitment may not be uttered; (iii) R_{quest}: The question P? can be answered only by P, 'Not P' or 'No commitment P'; (iv) R_{chall}: 'Why P?' must be responded to by a withdrawal of P, a statement not under challenge by its speaker, or a resolution demand of any of the commitments of the hearer, which immediately imply P; (v) $R_{resolve}$: Resolution demands may be made only if the hearer is committed to an immediately inconsistent conjunction of statement, or withdraws or challenges an immediate consequent of his commitments; (vi) $R_{resolution}$: a resolution demand must be followed by withdrawal of one of the offending conjuncts, or affirmation of the disputed consequent.

However, there are criticisms of certain dialogue rules provided by DC. For example, Woods and Walton [1982] and Walton [1984] argue that DC erroneously bans certain sequence of question begging, and Maudet and Moore [2001] argue that the rule $R_{repstat}$ may prevent one from answering questions in a preferred way. It is not clear whether there are more problems related to the set of rules, and whether DC can prevent fallacious argument and common errors is therefore called into question. A systematic study is therefore necessary to evaluate the set of DC rules. This issue is important because one of the main utilities of the debating system we wish to build is to develop students' critical thinking and debating skills, and teach students how to avoid fallacious argument and common errors in a contentious debate (Yuan et al. 2002). This requires that the dialogue model can correctly prevent fallacious argument and common errors when they occur during the course of debate.

3 Conversational Agents

Walton [1998] argues that formal systems of dialogue are not sharply enough focused on practical contexts of argument use that need to be studied in relation to the fallacies. A more practical approach might be to use a computational environment as the test-bed to study the proposed model (cf. Maudet and Moore [2001]). A suitable means, we argue, is to allow two computer systems to run with a proposed model in dialogue with each other and study the result, since there is then less human involvement and it is easy to control the experimental variables. Further, conversational simulation is also stressed by Amgoud et al. [2000] to be an important means to get empirical results about dialogue models and their behaviors. Given this, it is necessary to build two conversational agents that can engage in debate with each other via DC, and analyze the dialogue transcripts. A computational test-bed which enables two computational agents (referred to henceforth as Simon and Chris) to conduct debate with each other via DC, has been built by the authors using Java. The system architecture is shown in figure 1.

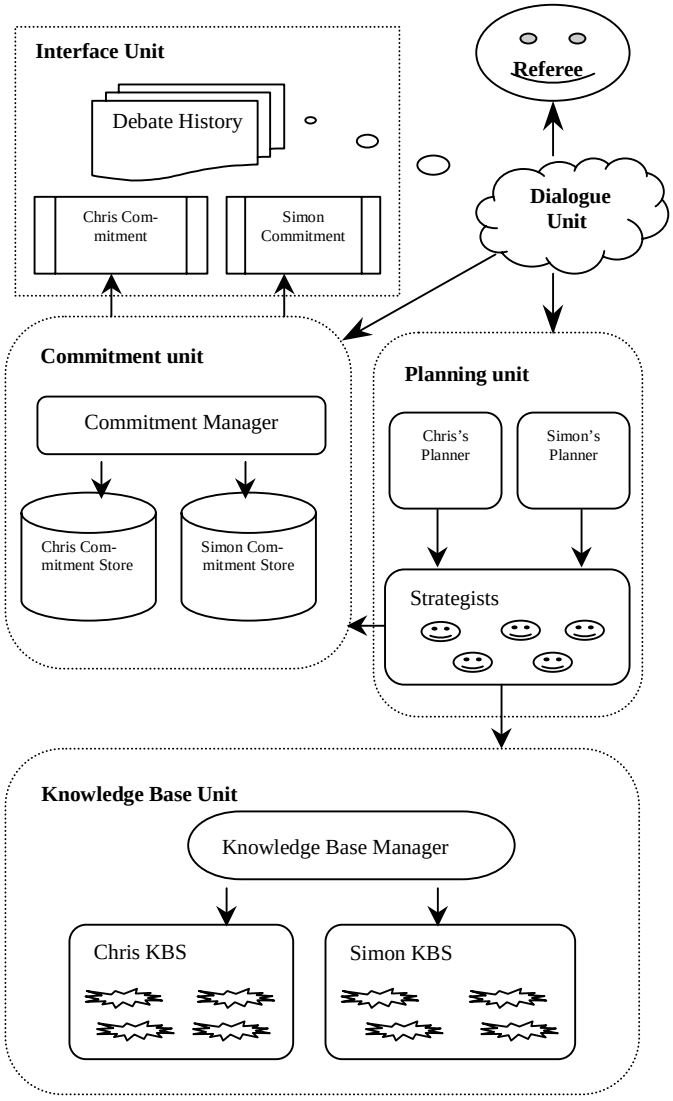

Figure 1. Conversational Agent System Architecture

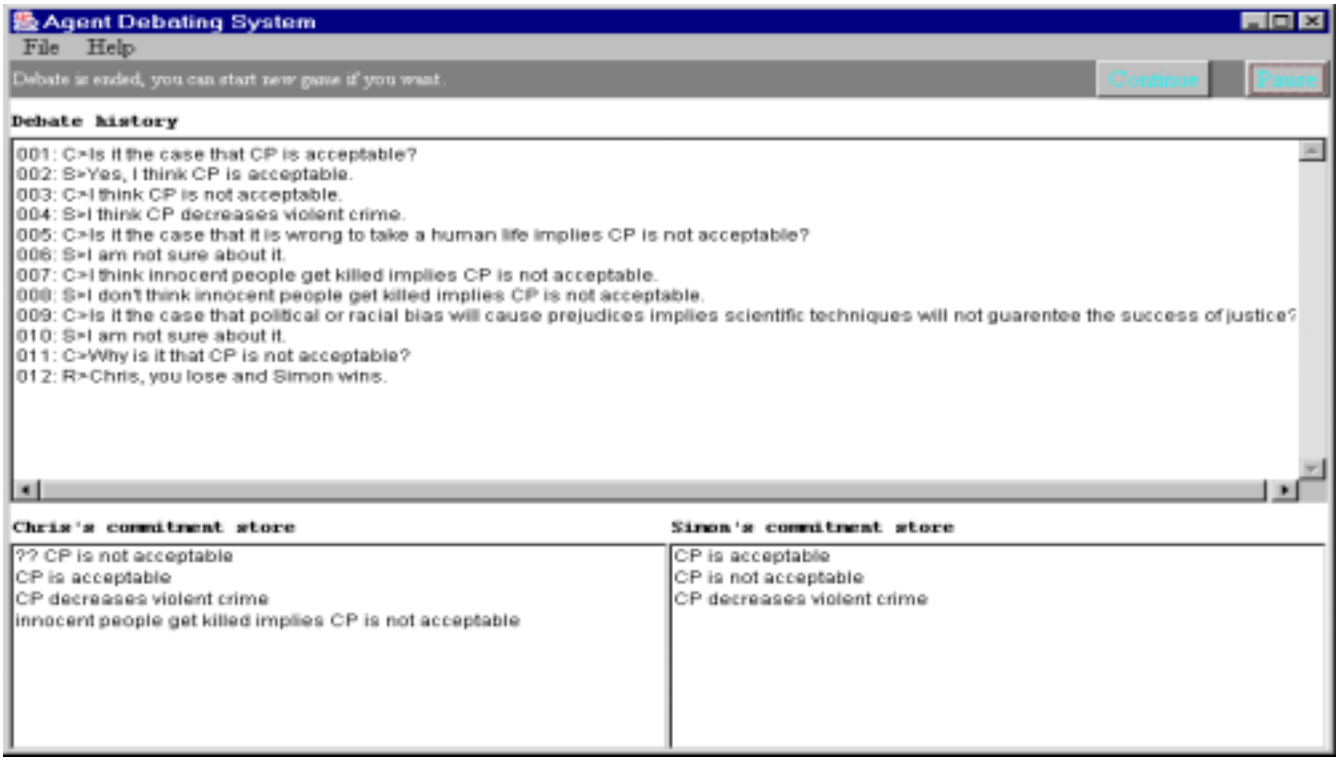

Figure 2. Interface of the Conversational Agent System

There are six main components of the system: interface unit, dialogue unit, referee, commitment unit, knowledge base unit and planning unit.

The interface unit provides a platform for agents' interaction and for the user to manipulate the interface. The system interface (see Figure 2) provides a dialogue history, which records the debate. The commitment-stores show both agents' commitment content. In order to control the process of the debate, a "New game" menu item is designed to start the debate, a "Pause" button is allocated to temporarily stop a debate, and a "Continue" button will carry on the dialogue. A "Save as" menu item is designed to save the dialogue history and both commitment sets as a separate file for subsequent analysis.

The dialogue unit can be regarded as the despatch centre of the agent interaction. This unit provides functions to upgrade the dialogue history and manage the turn taking of the agents according to the dialogue rule R_0. It will schedule the corresponding agent to make a move and then pass the move to the referee for judgement. If the move is legal, the commitment manager will be called to upgrade the commitment stores. Otherwise, the referee will promote a warning message and request the corresponding agent to make another move. Further, the role of the referee is to enforce the participants to follow the dialogue rule $R_{repstat}$ and detect whether each participant is using a statement under challenge to answer a challenge according to rule R_{chall}. The referee is also in charge of win or loss, the original DC regime makes no stipulation re winning and losing, but following Moore [1993], one agent will lose the debate when his thesis is removed from its store and the opponent's thesis is added (and not subsequently withdrawn) in its store.

The planning unit for an agent is responsible for generating moves according to the knowledge base of that agent and the prevailing state of both commitment stores. Each agent has its own planner but they share the same set of strategists. There are assertion, challenge, withdrawal, resolution and question strategists, which are designed to deal with different dialogue situations: (i) question strategist will generate three choices: Yes, No and I am not sure about it according to rule Rquest; (ii) challenge strategist will produce a set of moves according to rule Rchall, which consists of a resolution demand if the challenged statement is a consequence of partner's store, withdrawal of the statement being challenged, and assertion of any propositional statement in its own knowledge base. The job of enforcing participants to use statement not under challenge specified in rule Rchall is assigned to the referee; (iii) resolution strategist will produce a set of moves, which consists of withdrawal any of the conflicting conjuncts and affirmation of the disputed consequence according to rule Rresolution; *(iv)* DC has no restriction on the move type and move content after a statement or a withdrawal. The possible set of moves

produced by assertion and withdrawal strategists therefore consists of asserting, asking questions about any statement in its own knowledge base, requesting a resolution demand according to the rule Rresolve, withdrawal of any statement in its own commitment set, and challenge any propositional statement in partner's store. It is worth noting that these strategists will produce a pool of move choices, and the planner, in the current version of the system, will randomly select one and pass it to the dialogue unit to make a contribution. The purpose of using random arguments is to ensure that the dialogue transcripts generated by agents cover different aspects of the proposed model.

The commitment unit is responsible for upgrading agents' commitment stores. It includes a commitment manager and two commitment stores, one for each agent. The commitment manager will update both agents' commitment stores according to DC commitment rules. Each agent's commitment store is designed to have two lists of statements, the overall commitments of the party and statements under challenge. After each agent's move, both commitment stores will repaint the system interface, and any statement under challenge is marked with "??" shown on the interface (see figure 2).

The knowledge base unit consists of a knowledge base manager and both agents' knowledge bases. When the game starts, the dialogue manager will invoke the knowledge base manager to initialise both agents' knowledge bases. One agent will be set up to support the view of "capital punishment is acceptable", and the other to support "capital punishment is not acceptable". The knowledge base contains a set of propositions and consequence relationships between these propositions. The domain knowledge is formalised from Moore's [1993] experimental study of DC with human subjects.

4 Analyzing DC Transcripts

This section will discuss some of issues arising from the analysis of the agent-generated dialogue. Four sample dialogue transcripts will be discussed: DC1, DC2, DC3, DC4. The analysis is carried out in two ways: one is to detect whether there are fallacious argument if the two participants strictly follow DC regulations, the other is to analyse whether the referee's judgements are appropriate. These analysis show that in total participants made 258 (25+73+52+108) DC legal moves, and 19 (2+3+5+9) of them are found to be fallacious. DC referee makes 35 (5+10+8+12) judgements, and 24 (4+6+5+9) of them are analysed to be inappropriate. These suggest weakness in the DC rule set. In particular, these issues fall into the following categories.

4.1 Straw Man Fallacy

The straw man fallacy is a misrepresentation of someone's commitments in order to refute that person's argument [Walton 1996]. It is found from the transcripts that the use of resolution demand for an alleged inconsistency in some circumstances seem to involve an erroneous representation of the opponent's positions, e.g. the following extracts from DC2 (S for Simon, C for Chris).

001: C>Is it the case that CP is acceptable?
002: S>Yes, I think CP is acceptable.
003: C>I think CP is not acceptable.
004: S>Please resolve CP is acceptable and CP is not acceptable in your store.

It is shown from this dialogue fragment that Simon states his thesis "CP is acceptable" in turn 002, and Chris asserts the negation in turn 003. The alleged inconsistency of "CP is acceptable and CP is not acceptable" issued by Simon in turn 004 is obviously not real. Simon has misrepresented Chris's commitment and commits himself straw man fallacy. This reflects issues with DC commitment rules CR_S and CR_{YS}--statement will go to both commitment stores. This causes the statement "CP is acceptable" to go to Chris's store and as a result Chris is simultaneously committed to contrary statements. So, the commitment functions of DC may cause one to get into conflict purely because of statements made by the other party.

4.2 Restrictive Repetition

There are some occasions on which the DC referee is too restrictive when participants respond to questions. E.g. the following dialogue fragment drawn from DC1.

016: S>Is it the case that CP is acceptable?
017: C>Yes, I think CP is acceptable.
018: R>This statement is in both stores, you can not state it any more, please try again!
019: C>Yes, I think CP is acceptable.
020: R>This statement is in both stores, you can not state it any more, please try again!
021: C>No, I think CP is not acceptable.
022: R>This statement is in both stores, you can not state it any more, please try again!
023: C>Yes, I think CP is acceptable.
024: R>This statement is in both stores, you can not state it any more, please try again!
025: C>I am not sure about it.

The transcript shows that it is illegal for Chris to answer the question with "Yes" in turn 17,19 and 23, and with "No" in turn 21, the reason given by referee is that the statement is in both stores. The only choice for Chris is to withdraw the statement "CP is acceptable". Intuitively, this seems to give too much power to the questioner and should not be allowed in a fair dialogue. This issue revealed here is in line with Maudet and Moore's [2001] argument that the rule $R_{repstat}$ may prevent participants from answering questions in a preferred way.

Further, certain moves, which use a mutual commitment to answer a challenge might have been wrongly banned by DC referee as a repetition. We take the following dialogue fragment from DC4 as an example.

120: S>Why is it that 60% British people support CP?

121: C>Because murderers should not receive capital punish ment.
122: R>This statement is in both stores, you can not state it any more, please try again!

The dialogue fragment shows that Chris uses an agreed statement "murderers should not receive capital punishment" to answer a challenge in turn 121, but the referee raises the flag saying that the statement is in both stores. It looks intuitively unreasonable to prevent one using a previous agreed statement to answer a challenge. So, the rule $R_{repstat}$ may prevent one from answering the question in a preferred way and answering a challenge using an agreed statement.

4.3 Question Begging

Mackenzie describes question begging as "when one participant ask the other to grant him the premise which contains the substance of what is in dispute" [1979:117]. It is shown in the dialogue transcripts that certain sequences of moves are probably wrongly banned by DC as begging the question. We take the following dialogue fragment from DC4 as an example:

094: S>Why is it that innocent people get killed?
095: C>Because murderers should not receive capital punishment.
096: S>I don't think there are mistakes during judicial process implies innocent people get killed.
……
102: S>Why is it that scientific techniques will guarantee the success of justice?
103: C>Because innocent people get killed.
104: R>You may beg the question, please choose a state ment not under challenge and try again!

In the above dialogue fragment, Simon challenges the statement "innocent people get killed" in turn 094, which makes the statement under challenge with respect to Simon and means that Chris can not use it to answer a challenge any more, and this causes the DC referee judge that Chris is begging the question in turn 103. However, Chris has successfully defended the statement "innocent people get killed" in turn 095 and Simon accepts (or raises no objection to) it, and therefore the statement is not under dispute and Chris is not begging the question in turn 103. This evidence is actually in line with Woods and Walton's [1982] argument that DC erroneously bans certain sequence of moves as question begging.

Mackenzie later [1985; 1990; 1994] amended DC, via the substitution in clause (iii) of R_{chall} of the requirement that the ground statement be "acceptable", in place of the requirement that it be "not under challenge". He argues that the amended rule would cause only genuine question-begging sequences to be banned [1985: 335]. A statement S is acceptable to participant A at a stage n, just in case that either (i) S is a modus ponens consequence of A's store or (ii) S is not under challenge by A [1990: 575]. Fulfilling any one of the conditions can be judged as acceptable. It is true that this amendment will provide the service for the above dialogue fragment. In this case, the statement "innocent people get killed" is under challenge with respect to Simon, so the first condition for acceptable is not met. However, Simon de facto commits to it after turn 095, which means that the second condition of acceptable is met. Consequently, the statement "innocent people get killed" is acceptable to Simon and Chris will not be begging the question in turn 103.

However, Mackenzie's new amendment can be analyzed (though not implemented in current system) as still failing to ban certain sequences of what intuitively appear to be question-begging moves, e.g. the following extracts from DC3:

018: S>Why is it that truth is always on the side of most people?
019: C>Because truth is sometimes on the side of fewer people.
020: R>This statement is in both stores, you can not state it any more, please try again!
021: C>Because CP is not acceptable.
022: S>Is it the case that murderers should receive capital pun ishment?

In the above dialogue fragment, Simon does not commit to challenge of "CP is not acceptable" and therefore the first condition of acceptable is met, and Chris is not begging the question in turn 021 according Mackenzie's new amendment. However, it is intuitive that Chris is using his thesis as premise, which is under dispute to answer a challenge in turn 021 and is therefore begging the question. So, DC sometimes fails to deal with genuine question begging and erroneously bans certain sequences of moves as question begging. The issue of banning question begging therefore remains open, as is evidenced by Mackenzie's [1994] claim that none of Walton [1991] and his own theories of fallacies address the problem adequately.

4.4 Inappropriate Challenge

There are circumstances in the transcripts of a participant challenging a statement, which is not advanced by the opponent. For example, the following dialogue fragment is taken from DC2.

047: C>I think murderers should not receive capital punishment.
048: S>I think political or racial bias will not cause prejudice.
049: C>Please resolve CP is acceptable and CP is not acceptable in your store.
050: S>I don't think CP is not acceptable.
051: C>Why is it that murderers should not receive capital punishment?

In the above case, Chris claims that "murderers should not receive capital punishment" in turn 047, it is therefore strange that Chris challenges a statement made by himself in turn 051. In some occasions, participants are even challenging their own thesis, e.g. (DC1: 33; DC3: 66). Self-attacking might be seen as a poor strategic play rather than unfair protocol, but it reflects the issue that there is no pre-condition specified for a challenge in DC, which may cause one party to challenge a statement not advanced by both parties. This may be reasonable if restricted to information-

oriented dialogue, because participants can ask any information only if they do not know. But in a contentious debate, one may need to attack only the standpoints advanced by the other party (Van Eermeren et al.1996). So, the absense of a pre-condition for a challenge may lead to participants attacking a statement, which is not advanced by the other party.

5 Conclusions and Further Work

We have reported our work in using conversational software agents as a test bed to facilitate evaluation of the dialectical system DC in regulating an educational debate. Several issues of DC's preventing fallacious argument and common errors are brought to light by the analysis of agent-generated dialogues. This study provides further evidence to previous work concerning the argument of the weaknesses of DC in preventing fallacy of question begging (cf. Walton, [1984]), and appropriate handling of statement repetition (cf. Maudet and Moore [2001]. In addition, two possible weaknesses of DC in preventing straw man fallacy and the absence of a precondition for a challenge are revealed. In the light of these results, our immediate further work involves the design of a new dialogue game model. To make a convincing case that the new development does show improvement over DC in preventing fallacious argument and common errors, a similar conversational simulation and analysis will be conducted.

References

[Amgoud et al., 2000] Leila Amgoud, Nicolas Maudet and Simon Parsons. Modeling Dialogues Using Argumentation. *Proceedings of the Fourth International Conference on Multi-Agent Systems* (ICMAS-2000).

[Bench-Capon, 1998] Trevor J. M. Bench-Capon. Specification and Implementation of Toulmin Dialogue Game. *Proceedings of JURIX 98*, GNI, Nijmegen, pp. 5-20.

[Hamblin, 1970] Charles L. Hamblin. *Fallacies.* Methuen, London, 1970.

[Lodder and Herczog, 1995] Arno Lodder and Aimée Herczog. Dialaw, A Dialogue Framework for Modeling Legal Reasoning. *Proceedings of the fifth International Conference on Artificial Intelligence and Law* ACM, New York, pp. 146-155.

[Mackenzie, 1979] Jim Mackenzie. Begging the Question In Dialogue. *Journal of Philosophical Logic*, 8 (1979): 117-133.

[Mackenzie, 1985] Jim Mackenzie. No Logic Before Friday. *Synthesis* 63 (1985): 329-341.

[Mackenzie, 1990] Jim Mackenzie. Four Dialogue Systems. *Studia Logica* 49 (1990) 567-583.

[Mackenzie, 1994] Jim Mackenzie. Context of Begging the Question. *Argumentation* Vol. 8, pp.227-240.

[Maudet and Moore, 2001] Nicolas Maudet and David Moore. Dialogue games as Dialogue Models for Interacting with, and via, Computers. *Informal Logic* Vol. 21, No. 3 (2001): pp.219-243.

[Moore, 1993] David Moore. *Dialogue Game Theory for Intelligent Tutoring Systems.* Unpublished doctoral dissertation, Leeds Metropolitan University.

[Moore and Hobbs, 1996] David Moore and David Hobbs. Computational Use of Philosophical Dialogue Theories. *Informal Logic* Vol.18 No.2 pp131-163.

[Prakken, 2000] Henry Prakken. On Dialogue Systems with Speech Acts, Arguments, and Counterarguments. *JELIA 2000, LNAI 1919*, pp.224-238, 2000.

[Ravenscroft and Pilkington, 2000] Andrew Ravenscroft and Rachel Pilkington. Investigate by Design: Dialogue Models to Support Reasoning and Conceptual Change. *International Journal of Artificial Intelligence in Education* (2000), 11, 237-298.

[Van Eemeren et al., 1996] Frans H van Eemeren et al. *Fundamentals of Argumentation Theory, A Handbook of Historical Background and Contemporary Development.* Lawrence Erlbaum Associates Publishers.

[Veerman et al., 2002] Arja Veerman, Jerry Andriessen and Gellof Kanselaar. Collaborative Argumentation in Academic Education. *Instructional Science* 30: 155-186, 2002.

[Walton, 1984] Douglas Walton. *Logical Dialogue Games and Fallacies.* University Press of America.

[Walton, 1989] Douglas Walton. *Question-Reply Argumentation.* Greenwood Press.

[Walton, 1991] Douglas Walton. *Begging the Question: Circular Reasoning as a Tactic of Argumentation*. New York, Greenwood Press, 1991.

[Walton, 1996] The Straw Man Fallacy. Logic and Argumentation, Amsterdam, Royal Netherlands Academy of Arts and Science, North-Holland, pp.115-128.

[Walton, 1998] Douglas Walton. *The New Dialectics: Conversational Context of Argument*. University of Toronto Press.

[Walton and Krabbe, 1995]. Douglas Walton and Eric Krabbe. *Commitment in Dialogue: Basic Concept of Interpersonal Reasoning.* Albany NY: State University of New York Press.

[Woods and Walton, 1982] John Woods and Douglas Walton. Question Begging and Cumulativeness in Dialectical Games. *Nous* 16, pp. 585-605.

[Yuan et al., 2002] Tangming Yuan, David Moore and Alec Grierson. Educational Human Computer Debate, A computational Dialectics Approach. *Workshop on Computational Models of Natural Argument. ECAI 2002.*