Computational Agents as a Test-Bed to Study the Philosophical Dialogue Model "DE": A Development of Mackenzie's DC

TANGMING YUAN, *Leeds Metropolitan University*
DAVID MOORE AND
ALEC GRIERSON

Abstract: This paper reports research concerning a suitable dialogue model for human-computer debate. In particular, we consider the adoption of Moore's (1993) utilization of Mackenzie's (1979) game DC, means of using computational agents as the test-bed to facilitate evaluation of the proposed model, and means of using the evaluation results as motivation to further develop a dialogue model, which can prevent fallacious argument 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.

Résumé: Nous faisons un compte rendu de la recherche sur un modèle dialogique approprié pour les débats humains-ordinateurs. Nous examinons l'adoption de l'usage que Moore (1993) fait du jeux DC de Mackenzie (1979); nous étudions aussi les moyens d'utiliser des agents calculateurs pour faciliter l'évaluation des modèles proposés, les moyens d'employer les résultats tirés des ces évaluations pour encourager un plus grand développement du modèle dialogique, qui peut éviter des arguments fallacieux et des erreurs courantes. Nous anticipons que ce travail contribuera à l'évolution du dialogue humain-ordinateur, et aidera à éclairer différents sujets dans la dialectique.

Keywords: computational dialectics, human computer debate, DC, computational agents, dialogue games, DE, fallacy

1. Introduction

There is an increasing use of a computational dialectics approach in the area of human computer interaction (e.g. Grasso et al., 2000), agent communication (e.g. Reed, 1998; Jennings, 2001), mediation of legal reasoning (e.g. Bench-Capon et al. 2000), and artificial intelligence in general (Walton 2000). Previous papers (Maudet and Moore, 2001; Yuan et al., 2002) considered the use of a computational dialectics 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 steps. Firstly, we rehearse the argument for the adoption of Moore's (1993) utili-

zation of Mackenzie's (1979) game DC, and provide a brief introduction to the model. Secondly, we discuss means of evaluating the proposed model and construct two computational agents able to debate with each other via the proposed model. Thirdly, we analyze the dialogue transcripts generated by these agents as they use a DC-based dialogue model and we use the analysis to categorize problems concerning DC rules. We then use the results as motivation to propose a further system "DE". Finally, we conduct a similar conversational simulation and analysis of DE and argue that DE does show improvement over DC in preventing fallacious argument and common dialogue errors.

2. Moore's Utilization of DC

Following Moore (1993), which in turn derives from Walton (1989), it can be argued that a set of criteria is required for a suitable underlying model for a human-computer debate:

- the model needs to be persuasion style and symmetric in nature;
- the set of move types provided by the model should be adequate for expression;
- the model should leave enough room for strategic formation;
- the model should be able to prevent fallacious argument;
- the model should impose only a light cognitive load on the user;
- the model should be computationally tractable.

Using these criteria, we have conducted a comparative study of the most recent developments of dialogue models (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 rather than debate oriented. Prakken's and Lodder and Herczog's systems lack a question move type, and this may prevent students from asking the tutor questions, and tutors from questioning the student's understandings; this is undesirable from an 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 for the specific requirements of educational human computer debate. However, Moore's (1993) utilization of 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, the dialogue game DC is chosen as the base system for further study.

The formal argumentation system DC was developed by Mackenzie (1979) to study the fallacy of question begging, and later on amended by Moore (1993) as the underlying model for a potential intelligent tutorial system. The DC system is set up with two participants in dialogue with each other. Each party is assigned a commitment store, which records participants' commitments at stages of dialogue. Commitments can be added or deleted from commitment stores during the evolving dialogue. Participants' moves are regulated by a set of rules, which prohibits illegal events. The amended DC system can be outlined as follows (cf. Moore and Hobbs, 1996).

Available move types

1) Statements. P, Q, etc. and the truth-functional compounds of statements: "Not P," "If P then Q," "P and Q."

2) Questions. The question of the statement P is "Is it the case that P?"

3) Challenges. The challenge of the statement P is "Why is it supposed that P?" (or briefly, "Why P?").

4) Withdrawals. The withdrawal of the statement P is "No commitment P."

5) Resolution demands. The resolution demand of the statement P is "resolve whether P."

Commitment rules

1) Initial commitment, CR0: The initial commitment of each participant is null.

2) Withdrawals, CRW: After the withdrawal of P, the statement P is not included in the speaker's store.

3) Statements, CRS: A statement P results in P being added to each store.

4) Defense, CRYS: After a statement P, if the precedent event was "Why Q?", P and "If P then Q" are included in each store.

5) Challenges, 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.

Dialogue rules

1) RFORM: Participants may utter individual permitted locutions in turn.

2) RREPSTAT: Mutual commitment may not be uttered.

3) RQUEST: The question P may be answered only by P, "Not P" or "No commitment P."

4) RCHALL: "Why P?" must be responded to by a withdrawal of P, a statement not under challenge by the challenger, or a resolution

demand of any of the commitments of the hearer, which immedi ately imply P.

5) RRESOLVE: Resolution demands may be made only if the hearer is committed to an immediately inconsistent conjunction of statements, or withdraws or challenges an immediate consequent of his commitments.

6) RRESOLUTION: A resolution demand must be followed by withdrawal of one of the offending conjuncts, or affirmation of the disputed consequent.

However, there have been criticisms of certain dialogue rules provided by DC. For example, Woods and Walton (1982) and Walton (1984) argue that DC erroneously bans certain sequences of question begging, and Maudet and Moore (2001) argue that the rule RREPSTAT 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 the issue of whether DC can prevent all fallacious argument remains open. This issue is important because one of the main aims of our potential debating system is to develop students' critical thinking and debating skills, and to teach students how to avoid fallacious argument and common errors in a contentious debate (Yuan et al., 2002). Given this, the dialogue model should be able to pick out fallacious argument and common errors when they occur during the course of debate. The focus of this paper therefore concerns a systematic study of the set of DC rules via computational agents, to establish the extent to which it might fulfill one of the requirements of educational human computer debate, which is that the model can prevent fallacious argument and common dialogue errors.

3. DC Computational 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. Maudet and Moore (2001) further argue that a computational environment may be used as the test-bed to study proposed dialogue models. 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 are no human involvements and it is therefore 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 structure and its behaviors. Given this, it is necessary to build two computational agents that can debate with each other via DC, and analyze the dialogue transcripts. A computational test-bed that 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 the Java program language. The system architecture is shown in Figure 1.
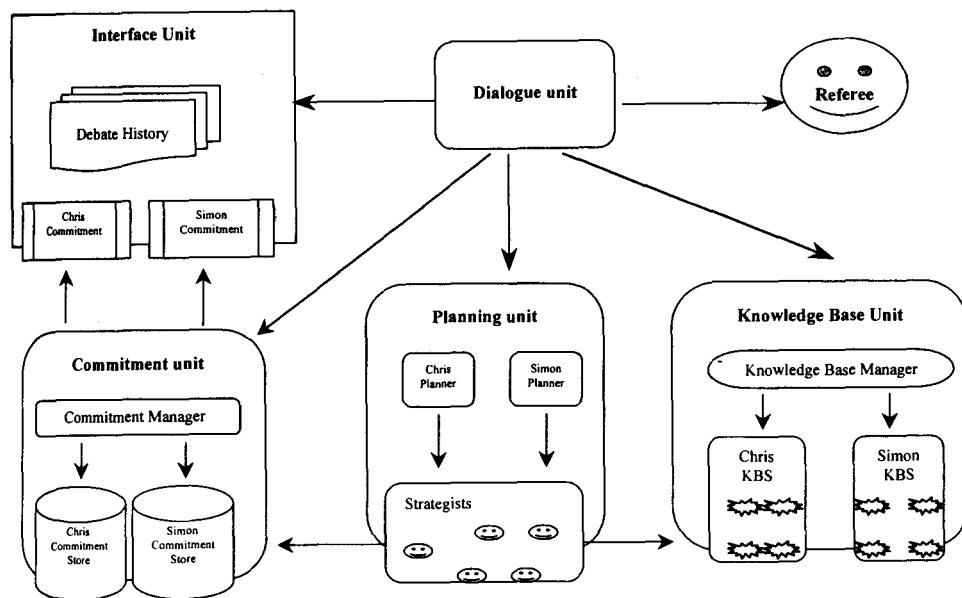
Figure 1. System architecture

There are six main components of the system. The interface unit (see Figure 2 which depicts two agents (Chris and Simon) debating capital punishment (CP)) provides a dialogue history, which will record the debate. The commitment-stores show both agents' commitment contents. In order to control the process of the debate, a "New game" menu item is designed to start the debate, a "Pause" button is available to temporarily stop a debate, and a "Continue" button will carry on the dialogue if necessary. 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 update the dialogue history and manage the turn taking of the agents and referee. This unit 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 update the commitment stores. Otherwise, the referee will post a warning message and request the corresponding agent to redo its move. Thus, the role of the referee is to enforce the participants to follow the DC dialogue rules. In addition, the referee is also in charge of win or loss; the original DC regime makes no stipulation regarding

winning and losing, but following Moore (1993), one agent will lose the debate when its thesis is removed from its store and the opponent's thesis is added to (and not subsequently withdrawn from) its store.
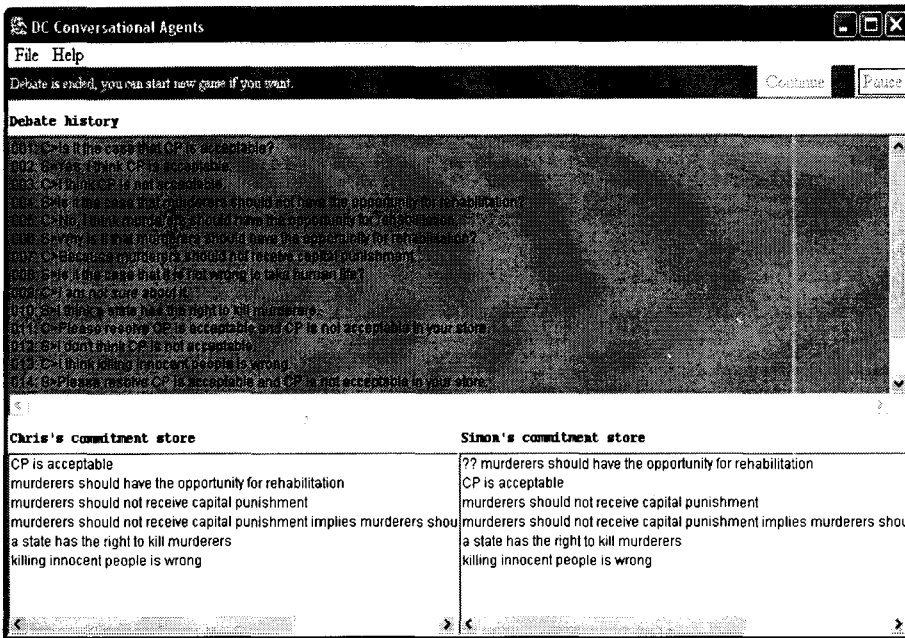

Figure 2. DC agents interface

The planning unit for an agent is responsible for generating moves in the light of 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 deals with incoming questions and will generate three choices: "Yes", "No" and "I am not sure about it" according to rule RQUEST. (ii) Challenge strategist deals with incoming challenges and will produce a set of moves, which consists of a resolution demand if the challenged statement is a consequence of partner's store, withdrawal of the statement being challenged, and asserting any propositional statement in its own knowledge base, according to rule RCHALL. (iii) Resolution strategist handles resolution requests, and it will produce a set of moves, which consists of withdrawal of any of the conflicting conjuncts and affirmation of the disputed consequence according to rule RRESOLUTION. (iv) Assertion and withdraw strategists are designed to handle the situation of a statement or a with-

drawal, they will produce a set of moves which consists of asserting or 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 of any statement in partner's store. In the current system, these strategists produce a pool of move-choices. The planner then selects one (currently on a random basis) and passes it to the dialogue unit to make a contribution. Using random argument may result in the dialogue transcripts generated by agents covering different aspects of the proposed model. It is anticipated that this will reveal certain failures of the dialogue rules that might be overlooked by manual use of the rules.

The commitment unit is responsible for maintaining 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. After each agent's move, both commitment stores will repaint the system interface; any statement under challenge is marked with a "??" as shown in Figure 2.

The knowledge base unit consists of a knowledge base manager and the knowledge bases of the two agents. 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 that "capital punishment (CP) is acceptable", and the other to support "capital punishment is not acceptable". The domain knowledge is formalised from Moore's (1993) experimental study of DC with human participants. The agent knowledge bases contain a set of propositions and the consequence relationships between these propositions.

4. Analyzing DC Transcripts

This section will discuss some of issues arising from the analysis of the agent-generated dialogue. Extracts from four dialogue transcripts: DC1, DC2, DC3, DC4 will be discussed (full transcripts are located at *http://www.geocities.com/ yuantangming/agentdialoguetranscript/DCDialogueTranscript.doc*). The analysis is carried out in two ways: one is to detect whether there are fallacious arguments if the two participants strictly follow DC regulations, the other is to analyse whether the referee's judgements are appropriate. These analyses 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 made 35 (5+10+8+12) judgements, and 24 (4+6+5+9) of them are analysed to be inappropriate in that DC wrongly bans certain sequences of moves as fallacious argument or dialogue errors. These figures suggest weaknesses 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 certain resolution demands for an alleged inconsistency might involve an intuitively erroneous representation of the opponent's positions, e.g. the following extracts from DC2 (S for Simon, C for Chris, R for referee).

001: C>Is it the case that CP is acceptable?   *Move type: question*

002: S>Yes, I think CP is acceptable.   *Move type: statement;*
rule: CRS, RQUEST

003: C>I think CP is not acceptable.   *Move type: statement;*
rule: CRS

004: S>Please resolve CP is acceptable   *Move type: resolution*
and CP is not acceptable in your store.   *demand; rule:*
RRESOLVE

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. The DC rules commit Chris to an inconsistency, which does not really represent his true position. Thus, when Simon attacks this position via a resolution demand, a form of straw man fallacy is, we submit, being committed. This reflects issues with DC commitment rules CRS and CRYS—a 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 contradictory statements. The two-way commitment function of DC may, that is, 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 intuitively too restrictive when participants respond to questions. Consider, the following dialogue fragment drawn from DC1.

016: S> Is it the case that CP is acceptable?   *Move type: question*

017: C>Yes, I think CP is acceptable.   *Move type: statement;*
Rule: RQUEST

018: R>This statement is in both stores,   *Rule: RREPSTAT*
you can not state it any more, please try again!

019: C>Yes, I think CP is acceptable.   *Move type: statement;*
Rule: RQUEST

020: R>This statement is in both stores,   *Rule: RREPSTAT*
you can not state it any more, please try again!

021: C>No, I think CP is not acceptable. *Move type: statement; rule: RQUEST*

022: R>This statement is in both stores, *Rule:RREPSTAT*
you can not state it any more, please try again!

023: C>Yes, I think CP is acceptable. *Move type: statement; rule: RQUEST*

024: R>This statement is in both stores, *Rule: RREPSTAT*
you can not state it any more, please try again!

025: C>I am not sure about it. *Move type: withdrawal; rule: CRW, RQUEST*

The transcript shows that it is illegal for Chris to answer the question with "Yes" in turn 017, 019 and 023, and with "No" in turn 021, the reason given by referee is that the statement is in both stores. The only choice for Chris is therefore 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 RREPSTAT 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 the 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 *Move type: challenge; rule: CRY*
support CP?

121: C>Because murderers should not *Move type: statement; rule: RCHALL*
receive capital punishment.

122: R>This statement is in both stores, *Rule: RREPSTAT*
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 previously agreed statement to answer a challenge.

4.3 Begging the Question

Mackenzie describes begging the question "when one participant asks the other to grant him the premise which contains the substance of what is under dispute" (1979, p. 117). It is shown in the dialogue transcripts that certain sequences of moves are arguably 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?	*Move type: challenge; rule: CRY*
095: C>Because murderers should not receive capital punishment.	*Move type: statement; rule: RCHALL*
096: S>I don't think there are mistakes during judicial process implies innocent people get killed.	*Move type: withdrawal; rule: CRW*

....

102: S>Why is it that scientific techniques will guarantee the success of justice?	*Move type: challenge; rule: CRY*
103: C>Because innocent people get killed.	*Move type: statement*
104: R>You may not beg the question, please choose a statement not under challenge and try again!	*Rule: RCHALL*

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 to 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 this (or at least raises no objection to it). 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 sequences of moves as question-begging.

Mackenzie (1985; 1990; 1994) later amended DC, via the substitution in clause (iii) of RCHALL 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 either or both of the conditions can be judged as making S acceptable. It is true that this amendment will remove the problem illustrated in the above dialogue fragment. In this case, the statement "innocent people get killed" is under challenge with respect to Simon, so the second condition for "acceptable" is not met. However, Simon de facto commits to it after turn 095, which means that the first condition of "acceptable" is met. Consequently, the statement "innocent people get killed" is acceptable to Simon and Chris will therefore not be begging the question in turn 103.

However, Mackenzie's new amendment (which is not implemented in the current system) seems to fail to ban certain sequences of what intuitively appear to be question-begging moves, e.g. the following extract from DC3:

018: S>Why is it that truth is always on the side of most people?	*Move type: challenge; rule: CRY*
019: C>Because truth is sometimes the side of fewer people.	*Move type: statement;* on *rule: RCHALL*
020: R>This statement is in both stores, you can not state it any more, please try again!	*Rule: RREPSTAT*
021: C>Because CP is not acceptable.	*Move type: statement; rule: CRYS, RCHALL*
022: S>Is it the case that murderers should receive capital punishment?	*Move type: question*

In the above dialogue fragment, Simon does not commit to challenge of "CP is not acceptable" and therefore the second condition of "acceptable" is met, and Chris is not begging the question in turn 021 according to Mackenzie's new amendment. However, Chris is using his disputed thesis to answer a challenge in turn 021 and we argue is therefore begging the question.

The result seems to be that DC, even with the "acceptable" statement amendment, 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 instances in the transcripts of a participant challenging a statement, which has not been advanced by the opponent. For example, the following dialogue fragment is taken from DC2.

047: C>I think murderers should not receive capital punishment.	*Move type: statement; rule: CRS*
048: S>I think political or racial bias will not cause prejudice.	*Move type: statement; rule: CRS*
049: C>Please resolve CP is acceptable CP is not acceptable in your store.	*Move type: resolution* and *demand; rule: RRESOLVE*
050: S>I don't think CP is not acceptable.	*Move type: withdrawal; rule: CRW*
051: C>Why is it that murderers should not receive capital punishment?	*Move type: challenge; rule: CRY*

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. Indeed, on some occasions, participants are even challenging

their own thesis, e.g. (DC1: 033; DC3: 066). 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 the other party. This may be reasonable if restricted to information-oriented dialogue, because participants can ask for 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 Eemeren et al., 1996).

In sum, several issues concerning DC's ability to prevent fallacious argument and common errors are brought to light by the agent-generated dialogues. First, the commitment rules of CRS and CRYS may cause some unnecessary conflict and one may get into conflict purely because of statements made by the other party. Second, the rule RREPSTAT may prevent one from answering a question in the preferred way and from answering a challenge using an agreed statement. Third, DC sometimes fails to deal with genuine question begging and erroneously bans certain valid sequences of moves as question begging. Finally, the absence of a pre-condition for a challenge may lead to participants attacking a statement, which is not advanced by the other party. In the light of these results, a further system, DE, has been designed. This will be discussed in the next section.

5. System DE

System DE has the same set of move types as DC since no problem has been found concerning the set of move types as such, but the commitment rules and dialogue rules are modified as discussed below.

5.1 Commitment rules

Each party owns a commitment store, and each commitment store contains two lists of statements: the "assertion list" contains the statements the participants have stated and the "concession list" contains the statements the participants have implicitly accepted (i.e. against which they have raised no objection).

Statements, CRS: After a statement P, unless the preceding event was a
           challenge, P is included in the speaker's assertion list and
           hearer's concession list, and "Not P" will be removed from
           the speaker's concession list if it is there.

Defense, CRYS: After a statement P, if the preceding event was "Why Q?",
           "P" and "If P then Q" are included in the speaker's as-
           sertion list and the hearer's concession list. "Not P" and "Not (If
           P then Q)" are removed from the speaker's concession list if
           they are there.

Challenges, CRY: A challenge of P results in P being removed from the store
           of the move maker if it is there.

In DC, the commitments are put together into one list. However, intuitively these commitments may be appropriately divided into two different types: asserted statement and implicitly accepted statement. An assertion shows one's standpoint in the dialogue, a statement that the participant is obliged to defend if the other party requests him to do so, whereas a concession is a statement that the participant is committed to in a weak sense, which carries no burden of proof (Walton and Krabbe, 1995). Therefore the commitment store in DE is separated into two lists: assertions and concessions. The assertion list contains statements one has asserted; the concession list contains statements made by the other party that one has only implicitly accepted.

The rules CRO and CRW remain unchanged and need no comment. The clause "asserting a statement will remove the negation from the speaker's concession set" is added to the commitment rules CRS and CRYS. This addition will solve the problem, discussed earlier, of the unnecessary conflict caused purely by statements made by the other party.

The clause of DC's CRY "After the challenge of P, P is added to the hearer's store" is dropped on the grounds that one should be able to challenge only propositions to which one's dialogue partner is explicitly committed. The clause in the original CRY "the challenge 'Why P?' is included in the speaker's commitment" is dropped, because DE uses a different approach (see section 5.2 for details), rather than commitment to challenge, to tackle the issue of question-begging.

5.2 DE dialogues rules

The rules that differ from DC are as follows:

RREPSTAT: Mutual commitment may not be uttered unless to answer a question or a challenge.

RCHALL: "Why P?" must be responded to by a withdrawal of P, a statement acceptable to the challenger or a resolution demand of any of the commitments of the challenger, which immediately imply P.

RLEGALCHALL: "Why P?" may not be used unless P is on the assertion list of the hearer.

RFORM, RQUEST, RRESOLVE and RRESOLUTION remain unchanged from DC and need no comment. In order to address the issues of DC's unnecessary restriction on repeat statements, the rule RREPSTAT is modified to ban repeated voluntary statements only, and allow one to answer a question in the preferred way or answer a challenge using an agreed statement.

To address the issue of lacking pre-condition for a challenge in DC, one extra dialogue rule is added, namely RLEGALCHALL—"'Why P?' may not be used unless P is on the hearer's assertion list". One is allowed to attack only standpoints advanced by the other party (van Eemeren 1996), otherwise "Why P?" is academic (Woods and Walton, 1982, p. 590).

To address the issue that DC still fails to ban certain sequence of question begging, the dialogue rule RCHALL is modified as "acceptable" in place of the requirement that it be "not under challenge". It should be noted, however, that the concept of "acceptable" is different from Mackenzie's (1990, 1994) notion. Mackenzie sets two conditions for a statement S to be 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, p. 575). Mackenzie's first condition seems reasonable, because if A has de facto committed to S, then S should be acceptable to A. However, the second condition is problematic because the fact that A is not committed to "Why S?" does not necessarily imply that S is acceptable to A, as evidenced by the agent-generated dialogue analyzed in section 4.3. In the game DE, we propose a statement S to be acceptable to participant A at a stage n, just in case that S is (i) a commitment or (ii) a de facto commitment or (iii) a new commitment of A's store. Fulfilling one or more of the conditions can be judged as acceptable. The first condition, "S is a commitment", is intuitively obvious, and the second condition "a de facto commitment" is the same as Mackenzie's first condition. The third condition considers that a new commitment with respect to A's store is acceptable. A new commitment means a statement which has never been to A's store before that stage of dialogue. The rationale for a new commitment to be acceptable is to allow chaining of challenge in dialogue. The following example will help to understand how DE deals with the fallacy of begging the question.

Turn	Participants	Move	Simon's store	Chris's store
01	Simon	P	P	P
02	Chris	Why P?	P	
03	Simon	R	P, R, R→P	R, R→P
04	Chris	Why R?	P, R, R→P	R→P
05	Simon	P		

In the above dialogue, the commitment stores are updated according to DE commitment rules. Simon asserts P in turn 01 and P goes to both stores, Chris challenges P in turn 02 and this removes P from his store. Simon answers the challenge in turn 03 and R, R→P go to both stores. Again, Chris challenges R in turn 04 and this removes R from his store. The statement R issued by Simon in turn 03 has never been to Chris's store before that stage of dialogue, and it is therefore a new commitment with respect to Chris; and the third condition of "acceptable" is therefore met and Simon is not begging the question in turn 03. However, the statement P in turn 05 is neither a commitment nor a de facto commitment, nor a new commitment of Chris's store (since P has been to Chris's store after turn 01); neither condition is met and the statement P is therefore not acceptable to Chris and Simon is begging the question in turn 05.

6. DE Computational Agents

To investigate whether system DE does in practice show improvement over DC in preventing fallacious argument and common errors, a similar study has been carried out on the system DE. Two computational agents that can debate with each other via the game DE have been built. To ensure that only the independent variable (DC or DE) affects the generated dialogue transcripts, the system architecture, dialogue unit, knowledge base unit and planning unit remain unchanged from the DC system, but the commitment functions and dialogue rules are replaced by the DE regime. Indeed, the reusability of the computational agents system makes it a general tool to study any formal dialogue models.

The commitment manager updates both agents' commitment stores according to DE commitment rules. Each commitment store is designed to have two lists of statements, those that have been stated and those that have been implicitly accepted, and they are shown in one list on the interface. In order to distinguish them from each other, any statement that is only implicitly accepted is marked with a '*', as shown in Figure 3.
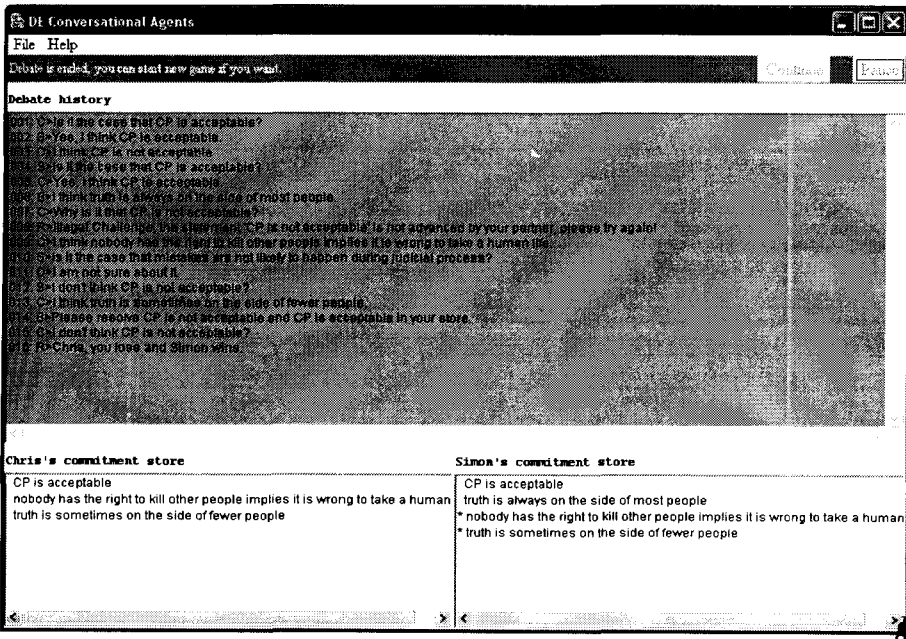

Figure 3. DE Agents Interface

The DE referee is responsible for enforcing participants to follow the DE rules including the new rules RREPSTAT and RLEGALCHALL, and for detecting whether

a participant is using an acceptable statement to answer a challenge according to the modified rule RCHALL.

7. Analyzing DE Transcripts

Three dialogue transcripts DE1, DE2 and DE3 have been selected (full transcripts are located at *http://www.geocities.com/yuantangming/ agentdialoguetranscript/DEDialogueTranscript.doc*). In total, participants made 146 (51+40+55) DE legal moves, DE referee made 10 (2+3+5) contributions. We first look at the fallacy of begging the question, the straw man fallacy, inappropriate challenge and statement repetition, to establish whether DE successfully deals with these issues.

Other fallacies such as complex question, appeals to emotions and irrelevant conclusions (cf. Walton, 1984; Woods et al., 2000) are then briefly discussed.

7.1 Question Begging

First, it is necessary to look at whether DE fails to ban certain sequences of question begging. There are 11 DE legal moves (DE1: 044, 057, 061; DE2: 010, 032, 041; DE3: 016, 018, 020, 048, 052), which are using a statement to answer a challenge. Among them, 7 instances (DE1: 044, 057, 061; DE2: 010, 032; DE3: 016, 020) of statements used are new commitments with respect to both commitment stores, they are not under dispute and therefore the speaker is not begging the question. On 3 occasions (DE2: 041, DE3: 018, 052) statements used are agreed (mutually committed) statements and hence the speaker again cannot be accused of begging the question. The remaining instance (DE3: 048) needs some discussion:

017: C>Why is it that scientific techniques will guarantee the success of justice?	*Move Type: challenge; rule: CRY*
018: S>Because murderers should receive capital punishment.	*Move type: statement; rule: CRYS, RCHALL*
....	
047: C>Why is it that 60% British people do not support CP?	*Move Type: challenge; rule: CRY*
048: S>Because scientific techniques will guarantee the success of justice.	*Move type: statement; rule: CRYS, RCHALL*
049: C>Why is it that repeat crime rate is extremely low?	*Move Type: challenge; rule: CRY*

In the above fragment, Simon is answering a challenge in turn 048 with the statement "scientific techniques will guarantee the success of justice" which Chris has de facto committed to after turn 018 and is therefore not under dispute. Simon is therefore not begging the question in turn 048. According to the discussions above, then, no fallacy of begging the question has been found from the DE legal moves.

A second issue is to see whether DE wrongly bans certain sequences of moves as question begging. There are three occasions (DE1: 055; DE2: 008; DE3: 050) that are judged by the DE referee as question-begging. These dialogue fragments are reproduced below.

Case 1 (DE1)

003: C>I think CP is not acceptable.   *Move type: statement;*
     rule: CRS

....

026: S>I don't think CP is not acceptable.   *Move type: with-*
     drawal; rule: CRW

....

054: S>Why is it that scientific techniques   *Move Type: challenge;*
will not guarantee the success of justice?   *rule: CRY*

055: C>Because CP is not acceptable.   *Move type: statement*

056: R>You are begging the question 'CP   *Rule: RCHALL*
is not acceptable' which is under dispute,
please try again!

Case 2 (DE2)

002: S>Yes, I think CP is acceptable.   *Move type: statement;*
     rule: CRS

003: C>I think CP is not acceptable.   *Move type: statement;*
     rule: CRS

....

007: C>Why is it that CP is a good deterrent?   *Move Type: challenge;*
     rule: CRY

008: S>Because CP is acceptable.   *Move type: statement*

009: R>You are begging the question 'CP is   *Rule: RCHALL*
acceptable' which is under dispute,
please try again!

Case 3 (DE3)

016: S>Because it is not wrong to take   *Move type: statement;*
human life.   *rule: CRYS, RCHALL*

....

039: C>I don't think it is not wrong to   *Move type: with-*
take human life.   *drawal; rule: CRW*

....

049: C>Why is it that repeat crime rate is extremely low?	*Move Type: challenge; rule: CRY*
050: S>Because it is not wrong to take human life.	*Move type: statement*
051: R>You are begging the question, 'it is not wrong to take human life' is under dispute, please try again!	*Rule: RCHALL*

Cases 1 and 2 show respectively that participant is using his thesis "CP is not acceptable", "CP is acceptable" as a premise in turn 055, 008; since these are intuitively under dispute at that stage, the participant is begging the question. Case 3 shows that Simon is using the statement "it is not wrong to take human life" as a premise in turn 050, a statement contested by Chris in turn 039, and hence is begging the question. Given the above analysis, DE referee's judgements on begging the question seem correct.

7.2 Straw man fallacy

To examine whether DE can prevent the straw man fallacy, we need to check whether the resolution demands in the transcripts are correctly made. There are three resolution demands used (DE1: 025, 062; DE2: 050), their situations are similar and we take an example from DE1:

Case 4 (DE1)

002: S>Yes, I think CP is acceptable.	*Move type: statement; rule: CRS*
003: C>I think CP is not acceptable.	*Move type: statement; rule: CRS*
....	
025: C>Please resolve CP is acceptable and CP is not acceptable in your store.	*Move type: resolution demand; rule: RRESOLVE*

The commitment policy is "silence means assent" (Mackenzie, 1979). For any statement made by one party, it will go to both commitment stores; if the hearer does not agree, he can withdraw, challenge or deny it and thus delete it from his store. In case 4, Simon states that "CP is acceptable" in turn 002, and it goes to Simon's assertion list and Chris's concession list, Chris does not agree with it and states "CP is not acceptable" in turn 003. According to DE's CRS, "CP is acceptable" will be removed from Chris's concession list and statement "CP is not acceptable" goes to Chris's assertion list and Simon's concession list. Now, Simon is simultaneously committed to "CP is acceptable" and "CP is not acceptable", however, Simon keeps silent without issuing a withdrawal, challenge or denial of the statement "CP is not acceptable", therefore the resolution demand is issued by

Chris in turn 025. The case of misrepresentation of opponent's position therefore still exists in DE. To avoid this, it can be suggested that a resolution demand can be made only on one's assertions rather than concessions. The rule RRESOLVE therefore requires further consideration.

7.3 Inappropriate challenge

To examine whether the DE referee correctly prevents one from challenging the standpoints not actually advanced by the partner, it is necessary to look at whether there are inappropriate challenge in the dialogue transcripts. There are 11 legal challenges (DE1: 043, 054, 060; DE2: 007, 031, 040; DE3: 015, 017, 019, 047, 049) found in the dialogue transcripts, and in each case the statements being challenged are advanced by the other party, suggesting that none of them are inappropriate. There are four occasions (DE2: 046; DE3: 024, 065, 067) that are judged by DE referee as illegal challenges and it is found in each case that the challenged statements is not made by the opponent, therefore we argue that the referee's judgements are appropriate.

7.4 Issue of repetition

Here we examine the appropriateness of the DE repetition regulation; this concerns repeated statements. There are three occasions on which repeated voluntary statements are banned by the referee (DE1: 014; DE2: 014; DE3: 030). Under these situations, the banned statement is a mutual commitment and therefore unnecessarily uttered again, so the referee's judgements are argued to be necessary.

7.5 Other issues

Other fallacies possibly occurring in dialectical systems such as complex question and appeals to emotions are not found from the dialogue transcripts. According to Woods et al. (2000, p. 61), the complex question fallacy occurs whenever a question is asked that contains a hidden, illicit, or unsupported assumption. In the current experimental setting, the propositions in the knowledge bases are carefully constructed avoiding two or more propositions rolled into one, and moves involving a conjunction of statements are not implemented in the current system. These arrangements may have already excluded the possibility of complex questions. The fallacy of emotions may include the abusive version of *ad hominem* and the boosterism version of *ad populum* and involves emotions going too far and hence interfering with the ultimate purpose of argument (Woods et al., 2000, p. 43). In the current system, the domain knowledge is carefully formalized to exclude strong emotional content, and this results in no fallacy of emotions being found in the agent-generated dialogue transcripts. The relationship between fallacies and knowledge representation is an interesting issue, which we leave for future work.

It might appear at first sight that the DE system permits the fallacy of irrelevant conclusions. According to Walton (1984), the fallacy involves giving evidence not proving the conclusion it was supposed to prove. Consider the following dialogue fragment from DE3.

017: C>Why is it that scientific techniques will guarantee the success of justice? *Move Type: challenge; rule: CRY*

018: S>Because murderers should receive capital punishment. *Move type: statement; rule: CRYS, RCHALL*

019: C>Why is it that CP is acceptable? *Move Type: challenge; rule: CRY*

In this dialogue fragment, the ground Simon is using in turn 018 is intuitively not relevant to the conclusion being argued for, yet the dialogue is legal according to DE. Furthermore, according to the DE (and indeed DC) commitment rule CRYS, after turn 18, "murderers should receive capital punishment" and "murderers should receive capital punishment *implies* scientific techniques will guarantee the success of justice" will go to both stores and hence commit both participants to the apparent irrelevance. However, Chris can challenge the conditional if he doubts the relevance of the argument and this suggests that the DE system does provide a means of participants avoiding being disadvantaged by irrelevant conclusions.

8. Conclusions and Further Work

We have used computational agents as a test bed to generate dialogue transcripts, which facilitates the evaluation of the dialectical system DC. Results have been used as motivation to construct a further system DE. DE is further evaluated using similar conversational simulations, and the result shows improvement over DC in preventing fallacious arguments and common errors. In particular, DE appears advantageous over DC in preventing the fallacy of question begging, inappropriate challenges and the straw man fallacy, and appropriate handling of the issue of repetition.

However, to be a suitable model for a human computer debating system, a further crucial issue concerns whether potential users accept and can successfully manipulate the dialogue model DE. This issue cannot be settled on an *a priori* basis. Empirical data is needed with regard to how far users become familiar with the system through exposure to it, ease of learning, awareness of a move's impact on the commitment stores and perceived satisfaction and comfort with the system. To study the user's acceptability of the proposed model, a crucial element is the existence of an actual implementation of the system in some form (Dix et al., 1998). Currently therefore a human-computer debate prototype is being developed using DE as the underlying model. However, the system's functionality will be crucially dependent upon appropriate strategic considerations (Moore, 1993, p. 185). A further issue therefore is to develop suitable computational strategies and incorporate them into the human-computer prototype. Further user-based studies will then be necessary to assess the effectiveness of the proposed strategy. This is the topic of on-going research.

References

Amgoud, L., Maudet, N., and Parsons, S. (2000). "Modeling dialogues using argumentation," *Proceedings of the Fourth International Conference on Multi-Agent Systems*.

Bench-Capon, T. J. M. (1998). "Specification and implementation of Toulmin dialogue game," *Proceedings of JURIX 98*, GNI, Nijmegen, pp. 5-20.

Bench-Capon, T. J. M., Geldard, T. and Leng, P. H. (2000). "A method for computational modeling of dialectical argument with dialogue games," *Artificial Intelligence and Law* 8, pp. 233-254.

Dix, A. J. et al. (1998). *Human-Computer Interaction.* (2nd ed.). Prentice Hall Europe.

Eemeren, F. H. van, et al. (1996). *Fundamentals of Argumentation Theory, A Handbook of Historical Background and Contemporary Development.* Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Grasso, F., Cawsey, A. and Jones, R. (2000). "Dialectical argumentation to solve conflicts in advice giving: a case study in the promotion of healthy nutrition," *International Journal of Human Computer Studies, vol.* 53, pp. 1077-1115.

Hamblin, C. (1970). *Fallacies.* London: Methuen.

Jennings, N. R. (2001). "An agent-based approach for building complex software systems," *Comms. of the ACM*, 44 (4), pp. 35-41.

Lodder, A. R. and Herczog, A. (1995). "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, J. (1979). "Begging the question in non-cumulative systems," *Journal of Philosophical Logic* 8, pp. 117-133.

Mackenzie, J. (1985). "No logic before Friday," *Synthesis* 63, pp. 329-341.

Mackenzie, J. (1990). "Four dialogue systems," *Studia Logica* 49, pp. 567-583.

Mackenzie, J. (1994). "Context of begging the question," *Argumentation,* 8, pp. 227-240.

Maudet, N. and Moore, D. (2001). "Dialogue games as dialogue models for interacting with, and via, computers," *Informal Logic* Vol. 21, No. 3 pp. 219-243.

Moore, D. (1993). *Dialogue Game Theory for Intelligent Tutoring Systems.* Unpublished Doctoral Dissertation, Leeds Metropolitan University.

Moore, D. J. and Hobbs, D. J. (1996). "Computational use of philosophical dialogue theories," *Informal Logic* Vol.18 No.2 pp. 131-163.

Prakken, H. (2000). "On dialogue systems with speech acts, arguments, and counterarguments," *JELIA 2000, LNAI 1919*, pp. 224-238.

Ravenscroft, A. and Pilkington, R. (2000). "Investigate by design: dialogue models to support reasoning and conceptual change," *International Journal of Artificial Intelligence in Education* 11, pp. 237-298.

Reed, C. A. (1998). "Dialogue frames in agent communication," in *Proceedings of the 3rd International Conference on Multi-Agent Systems (ICMAS98),* Paris: IEEE Press, pp. 246-253.

Veerman, A. et al. (2002). "Collaborative argumentation in academic education," *Instructional Science* 30, pp. 155-186.

Walton, D. (1984). *Logical Dialogue Games and Fallacies.* Lanham, MD: University Press of America.

Walton, D. (1989). *Question-Reply Argumentation.* New York: Greenwood Press.

Walton, D. (2000). "The place of dialogue theory in logic, computer science and communication studies," *Synthesis* 123, pp. 327-346.

Walton, D. and Krabbe, E. (1995). *Commitment in Dialogue: Basic Concept of Interpersonal Reasoning.* Albany NY: State University of New York Press.

Woods, J. and Walton, D. (1982). "Question begging and cumulativeness in dialectical games," *Nous* 16, pp. 585-605.

Woods, J., Irvine, A. and Walton, D. (2000). *Argument, Critical Thinking, Logic and the Fallacies.* Toronto: Pearson Education Canada.

Yuan, T., Moore, D. and Grierson, A. (2002). "Educational human computer debate, a computational dialectics approach," *Workshop on Computational Models of Natural Argument. ECAI 2002.*

Tangming Yuan, David Moore and Alec Grierson
School of Computing, Leeds Metropolitan University
Beckett Park, Leeds LS6 3QS, United Kingdom

Email: d.moore@leedsmet.ac.uk