A Geometric Semantics for Dialogue Game Protocols for Autonomous Agent Interactions

Peter McBurney and Simon Parsons
Department of Computer Science
University of Liverpool
Liverpool L69 7ZF U. K.
{p.j.mcburney,s.d.parsons}@csc.liv.ac.uk

February 4, 2002

Abstract

Formal dialogue games studied by philosophers since the time of Aristotle have recently found application in Artificial Intelligence as the basis for protocols for interactions between autonomous software agents. For instance, game protocols have been proposed for agent dialogues involving team formation, persuasion, negotiation and deliberation. There is yet, however, no formal, mathematical theory of dialogue game protocols with which to compare two protocols or to study their formal properties. In this paper, we present preliminary work towards such a theory, in which we develop a geometric semantics for these protocols and, with it, define a notion of equivalence between two protocols. We then demonstrate an algebraic property of protocol equivalence, and use this to show the non-equivalence of two similar generic protocols. We also explore the relationship between finite and infinite dialogues, motivated by the Ehrenfeucht-Fraïssé games of model theory. Our results have implications for the design and evaluation of agent dialogue-game protocols.

KEYWORDS: Agent Communication Languages, Argumentation, Computational Dialectics, Dialogue Games.

1 Introduction and Motivation

The concept of autonomous and intelligent software entities — *agents* — interacting to jointly achieve some objective has become an influential guiding principle for recent research in Artificial Intelligence (AI) [32]. A typical application involves agents representing divergent interests, e.g., different network elements in (say) a telecommunications or electricity network, seeking to negotiate use of a scarce resource. In these instances, economic auction mechanisms have provided powerful computational tools for operationalization of negotiation interactions between agents [15]. An auction mechanism, however, permits only limited information to be transmitted between its participants — typically, only acceptances or rejections of offers at particular price levels. Knowing the *reasons* for particular offers, and the *reasons* for or against their acceptance, can enable participants to refine their offers or to propose appropriate counter-offers. This additional information can thus potentially increase the likelihood of the parties reaching resolution, and, if achieved, possibly reduce the average time or computational resources required. For this reason, the use of argumentation, the philosophical theory of argument and dialogue, has recently been proposed for interactions between autonomous agents [26, 29].

How might we operationalize argumentation in the design of agent interaction protocols? One approach has been the use of formal dialogue games. These are interactions between two or more participants, where each "moves" by uttering a locution, i.e. a spoken utterance, in conformance with some pre-defined and commonly-known set of rules. A locution may, for example, be an assertion of some claim by a speaker, or a question regarding an assertion made by another speaker, or an argument for a claim in response to such a question. The rules of the game specify the circumstances under which each locution is permitted, or obligatory, or forbidden, and the circumstances under which the dialogue terminates. Dialogue games date at least from the time of Aristotle [3], and were extensively studied by philosophers in the Middle Ages [11]. In modern times, they have been used for the contextual analysis of fallacious reasoning [11, 18], to provide a semantics for intuitionistic and classical logic [17], and as a constructive proof-theory for statements in quantum logic [25]. They differ from the games of economic game theory in that payoffs for winning or losing a game are not considered, and because there is no use of uncertainty measures, such as probabilities, to model the possible moves of opponents. They also differ from the abstract games recently developed as a semantics for programming languages in theoretical computer science [1], since these latter games do not typically share the rich rule structure of dialogue games, and are not intended to be implemented as interaction protocols.

In the multi-agent systems arena, dialogue games have been proposed as the basis for protocols for several types of agent interactions, including: the formation of teams and collective intention [4, 5]; information-seeking dialogues [14]; persuasion dialogues [22]; negotiations [2, 24, 28]; joint deliberations over possible actions [12]; and joint determination of uncertain beliefs [22].[1] As an example, we list below the nine locutions in a protocol for deliberation over possible actions we have recently proposed [12]; this protocol implements a philosophical model of human deliberation, called *retroflexive argumentation*, due to Harald Wohlrapp [31].

[1] Note that some of these dialogue types are defined according to a standard argumentation theory classification of human dialogues [30]. However, the work discussed in this paper is unrelated to the use of dialogues games to explain or generate natural language conversations in computational linguistics, as in [16].

Locutions for a Deliberation Protocol

open_dialogue$(P_i, q?)$
enter_dialogue$(P_j, q?)$
propose(P_i, type, t)
assert(P_i, type, t)
prefer(P_i, a, b)
move$(P_i, \mathit{action}, a)$
retract$(P_i, \mathit{locution})$
ask_justify$(P_j, P_i, \mathit{type}, t)$
withdraw_dialogue$(P_j, q?)$

where: P_i, P_j *are agent identifiers;* $q?$ *is a question concerning what is to be done in some situation;* type *is one of the following categories of propositions:* question, goal, constraint, fact, perspective, action *or* evaluation*; and* t, a, b *are propositions* [12].

However, these various agent protocols are essentially ad-hoc, with no formal, unifying theory developed for them. Perhaps because of this, there has been as yet no study of the properties of such protocols, either particularly or in general, for example, whether they facilitate or hinder dialogue resolution; or whether they are vulnerable to disruptive behaviour by participants (e.g., by asking the same question repeatedly). Consequently, designers of dialogue game protocols currently have no guidance for issues such as:

- How many locutions should there be?
- What types of locutions should be included? (For example: assertions, questions, etc.)
- What are the appropriate rules for the combination of locutions?
- When should behavior be forbidden, e.g., repeated utterance of one locution?
- Under what conditions should dialogues be made to terminate?
- When are dialogues conducted according to a particular protocol guaranteed to terminate?
- For dialogues guaranteed to terminate, how many locutions need to be uttered for this to be achieved, i.e. what is the computational complexity of a dialogue under a given protocol?

Similarly, as protocols proliferate, agents (or their human principals) intending to use them will require guidance for issues such as:

- When are two protocols the same? Different protocols may be different only in superficial details, and not in the formal properties they exhibit; conversely, similar protocols may have very different properties.

- How may different protocols be compared and how may these differences be measured?
- Which protocols are to be preferred under which circumstances? (i.e., what are their respective advantages and disadvantages?)
- What is the formal relationship between dialogue game protocols and the various generic agent communications languages which have been proposed, e.g., the Knowledge Query and Manipulation Language (KQML) [19] or Agent Communications Language of the Foundation for Intelligent Physical Agents (FIPA) [9].
- When are dialogue game protocols preferable to other forms of agent interaction, such as auction mechanisms or general agent communications languages?
- What strategies are appropriate for an agent engaged in a dialogue under a specific protocol?

These are important questions for designers and users of agent dialogue game protocols. Answering them with rigor will require a formal language in which to reason about the properties of protocols, and techniques with which to compare one protocol with another. We believe that algebraic topological methods may have great value in developing such a mathematical theory of protocols, and this paper presents our preliminary investigations in this direction. Our approach is to define the elements of a generic dialogue game protocol in such a way that we can translate any protocol into a geometric representation, specifically a subset of a real, multi-dimensional space. We then define a notion of equivalence of two protocols, and construct an algebraic invariant of this notion of equivalence. Because dialogues correspond to time-directed paths in a euclidean space under our semantics, we have no notion of inverse, and so this algebraic invariant is not a group structure. Thus our work to date is in the spirit of algebraic topology without using anything sophisticated mathematically.

This paper is structured as follows: Section 2 reprises a formal model of dialogue games from our previous work, and Section 3 defines our geometric semantics for dialogue games protocols and for dialogues conducted according to such games. Section 4 defines a topological notion of equivalence of two protocols, and deduces an algebraic invariant of this notion. We then demonstrate that two generic protocols which differ by only one locution are in fact not equivalent. In Section 5 we explore the relationship between finite and infinite dialogues, using ideas motivated by the Ehrenfeucht-Fraïssé games of model theory [13]. Section 6 concludes the paper.

2 A Generic Dialogue-Game

Following our experiences in articulating the locutions and rules for a number of agent dialogue game protocols, e.g., [2, 22, 24], we proposed the first formal model for a generic dialogue game [20, 23]; we reprise this work here. In this model, it is assumed that the topics of discussion between the participants are represented in some logical language, whose well-formed formulae are denoted by the lower-case Greek letters, ϕ, θ, etc. The dialogue game protocol then includes several types of rules:

Commencement Rules: Rules which define the circumstances under which the dialogue begins.

Locutions: Rules which indicate what utterances are permitted. These rules commonly define the preconditions which must obtain before a locution can be uttered by a participant, the legal syntax of the locution, and the responses which other participants are permitted to utter subsequently. For instance, legal locutions may enable participants to assert propositions, allow others to question or contest prior assertions, and allow those asserting propositions which are subsequently questioned or contested to justify their assertions. Justifications may involve the presentation of a proof of the proposition or an argument for it, and such presentations may also be legal utterances. In some multi-agent system applications of dialogue games, e.g. [2]), sincerity conditions are imposed on utterances, for example allowing agents to assert statements only when they themselves have a prior argument or proof from their own knowledge base.

Combination Rules: Rules which define the dialogical contexts under which particular locutions are permitted or not, or obligatory or not. For instance, it may not be permitted for a participant to assert a proposition ϕ and subsequently the proposition $\neg\phi$ in the same dialogue, without in the interim having retracted the former assertion. Similarly, if one participant challenges a prior statement made by another participant, that latter participant may be required to defend the prior statement with specific locutions.[2]

Commitments: Rules which define the circumstances under which participants express commitment to a proposition. Typically, assertion of a claim ϕ in the debate is defined as indicating to the other participants some degree of willingness to defend it in the dialogue if questioned or contested by other participants. Since [11], it is common to track commitments in a set of publicly-readable blackboards called Commitment Stores.

Termination Rules: Rules which define the circumstances under which the dialogue ends. These rules may be expressible in terms of the contents of the Commitment Stores of one or more participants; for example, a persuasion dialogue may terminate when all participants utter a locution

[2] As an example of a combination rule in the deliberation dialogue protocol presented in the previous section, the locution **prefer**(P_i, a, b) cannot be uttered unless previous speakers have already uttered two or more locutions in which evaluations are made of the action-options a and b, i.e., in which a and b are referenced in propositions e and f respectively in locutions **assert**$(P_j, \mathit{evaluation},\ e)$ and **assert**$(P_k, \mathit{evaluation},\ f)$.

accepting the proposition at issue, a locution which inserts the proposition into the agent's Commitment Store.

In the dialogue game tradition in philosophy, *commitment* is a dialogical concept [11]: a commitment expresses a willingness by an agent to defend in the dialogue a claim which that agent has previously asserted, and does not necessarily have any connection to any reality external to the dialogue. Thus, for example, assertion of some proposition in a dialogue by a participant does not necessarily mean that this participant believes the asserted proposition outside the dialogue.[3] Dialogue locutions will only have connections with external reality if the participants in the dialogue together agree to vest such meanings in the locutions. However, for dialogues of interest to agent designers, particularly negotiations and deliberations, connections between locutions and external reality is important. For example, parties to a negotiation dialogue may agree that its successful conclusion will lead to the execution of a purchase transaction following dialogue termination. We therefore assume that participants to a dialogue may utter locutions which all participants understand to imply a willingness to execute an action or actions, external and subsequent to the dialogue.[4] We further assume that these actions may be expressed in the same logical language as are the topics of discussion.

3 A Geometric Semantics

In this section, we develop a geometric semantics for the framework just presented. We use $\mathcal{R}^+$ and $\mathcal{Z}^+$ to denote spaces consisting of all non-negative real numbers and integers, respectively. We use the notation $\mathcal{R}^{n+}$ for the n-fold product of $\mathcal{R}^+$, i.e. the sub-space of $\mathcal{R}^n$ where all points have non-negative co-ordinates on each dimension.

3.1 Dialectical systems

We assume a finite set $\mathcal{A} = \{P_i | i = 1, \ldots, p\}$, of dialogue participants, or agents. We denote dialogue games by possibly-subscripted upper case script Roman letters, $\mathcal{D}, \mathcal{E}$, etc. Each dialogue game comprises a finite set of legal locution-types, denoted $\mathcal{L} = \{L_j | j = 1, \ldots, l\}$, and a number of rules which we will consider presently. Dialogues conducted according to the rules of a dialogue game are assumed to concern a fixed, finite set $\Phi = \{\phi_i | i = 1, \ldots, q\}$ of well-formed formulae in some propositional language, which we call the universe of

[3] In [30, p. 72], such commitments are called *propositional*.

[4] In [23], we formalize these two notions of commitment, dialogical and semantic. Since a sufficiently clever agent can always simulate insincerely any required mental state, the problem of semantic verification of agent communications languages – how may we verify that all participants have the same sincere understanding of utterances? – is a thorny one, which we do not discuss here. For further details, see: [33]. As an example, Durrenmatt's novel *"Die Panne"* [6] concerns a human conversation to which the participants give very different meanings, with tragic consequences.

discourse.[5] Some or all of these propositions may represent external actions which can be the subject of commitments incurred by speakers in the dialogue uttering specific locutions, and so we distinguish a specific subset $\Phi_a \subseteq \Phi$, which we call the action set.

We further assume that all locution types can be categorized into one or more of three types: (a) *Commitment locutions:* locutions which express some external or dialogical commitment, e.g. a commitment to execute a purchase action upon completion of the dialogue; these locutions are instantiated with one or more elements from Φ_a and possibly one or more elements of $\Phi \setminus \Phi_a$; (b) *Information locutions:* locutions which transmit some information from speaker to audience, e.g. a statement by a speaker of a preference-ordering over some set of objects; these locutions are instantiated with one or more elements of $\Phi \setminus \Phi_a$; and (c) *Procedural locutions:* locutions which neither transmit information, nor commit to actions; such locutions may, for instance, encode a theory of argumentation, e.g. questions to or challenges of other speakers, etc; these locutions are not directly instantiated with elements of Φ, although they may be instantiated with other locutions in which elements of Φ appear.

We assume that utterances of dialogue locutions occur at discrete time-points, called *rounds*, represented by the non-negative integers. A locution is executed by a speaker P_i instantiating the appropriate locution type L_j with: a positive integer representing the time of utterance (a time-stamp); the identifier P_i of the speaker of the locution; and possibly a proposition ϕ_k from the set Φ.[6] We thus denote a locution L_j instantiated in this way by: $L_j(t, P_i, \phi_k)$. The actions referred to by instantiated utterances of Commitment locutions and the information transmitted by instantiated utterances of Information locutions are assumed to be expressible in formulae contained in the set Φ. Because some locutions may require instantiation with more than one proposition, e.g. a locution expressing a preference ordering over two purchase options in a negotiation, we assume Φ is closed under the combinations of propositions required by the syntax of the locution-types. We denote by $\mathcal{L}^*$ the finite set of locution-types $\mathcal{L}$ instantiated with possible speakers from $\mathcal{A}$ and discussion topics from Φ. In other words, $\mathcal{L}^*$ is isomorphic to a subset of the finite set of 3-tuples, $\{\ (\mathcal{L}, \mathcal{A}, \Phi)\ \}$.

In the previous section we identified the different types of rules in our model of a dialogue game. Because initiation of a dialogue of specific type on a specific topic happens outside it, we do not assume commencement rules are part of the definition of the dialogue game protocol. For all other rules, however, it is possible to define each rule as a mapping from appropriate dialogue histories (the instantiated locutions actually uttered prior to the next round) to the set $\mathcal{L}^*$ in the case of Combination and Termination rules, or to the set Φ, in the

[5] Of course, such a propositional language may have an infinite set of well-formed formulae. We leave discussion of the infinite case to another occasion.

[6] Some dialogue-game protocols, e.g., [24], permit speakers to target utterances at particular audiences, in which case the locution-type would also be instantiated with a subset of $\mathcal{A}$. For simplicity in this paper, we assume all utterances are intended for, and heard by, all participants.

case of Commitment rules. Details of these definitions can be found in [23]. For present purposes, it is sufficient to note that a dialogue game protocol includes a specification of such mappings, and that they induce a partition of the possible instantiated locutions for each agent at each round: *Obligatory moves:* instantiated locutions, one of which must be uttered by the agent at the next round; *Legal moves:* instantiated locutions which may be uttered by the agent at the next round; *Forbidden moves:* instantiated locutions which may not be uttered by the agent at the next round; and *Termination moves:* instantiated locutions which if uttered by the agent at the next round will result in termination of the dialogue. We next define a *dialectical system.*

Definition 1: *Suppose $\mathcal{A}$ is a finite set of agents, Φ a finite set of topics of discussion (including a possible subset of action commitments), and $\mathcal{D}$ a dialogue game protocol with a set of locutions $\mathcal{L}$. We say that the 4-tuple $\tilde{\mathcal{D}} = (\mathcal{A}, \Phi, \mathcal{L}, \mathcal{D})$ is a* dialectical system. *If the size of $\mathcal{A}$ is p, the size of Φ is q and the size of $\mathcal{L}$ is l, we say that $\tilde{\mathcal{D}}$ has dimension $n = pql$. A* dialogue *undertaken in accordance with such a dialectical system is a time-ordered sequence of locutions uttered by the agents in $\mathcal{A}$, each element of which consists of an instantiated locution from the set $\mathcal{L}^*$, uttered in accordance with the rules of dialogue game $\mathcal{D}$.*

We refer to dialogues undertaken in accordance with a dialectical system as being *associated with* or *under* the dialectical system. We also refer to such dialogues as *legal* dialogues, although we do not permit sequences of locutions under a dialectical system which do not conform to the rules of the corresponding dialogue game. In the next section, we present a geometric semantics for dialogue systems and their associated dialogues.

3.2 Dialogue paths

Let $\tilde{\mathcal{D}} = (\mathcal{A}, \Phi, \mathcal{L}, \mathcal{D})$ be a dialectical system, and set $n = pql$, where these constants are defined as in Section 3.1. We interpret the associated dialogues as paths in the real n-dimensional space $\mathcal{R}^{n+}$. We do this by labeling each axis of $\mathcal{R}^{n+}$ with a triple (P_i, ϕ_k, L_j), for all $P_i \in \mathcal{A}$, for all $\phi_k \in \Phi$ and for all $L_j \in \mathcal{L}$. The path corresponding to a dialogue, which we call a *dialogue path*, commences from the origin, and proceeds as follows: Whenever participant P_i utters

locution L_j concerning topic ϕ_k, the path moves from whatever is its current position forward one unit in a direction parallel to the axis labeled (P_i, ϕ_k, L_j). We first define such paths formally as follows:

Definition 2: *A* dialogue path *is a function $d(.) : \mathcal{R}^+ \rightarrow \mathcal{R}^{n+}$, such that conditions (a), (b) and (c) are each satisfied:*

(a) $d(0) = \tilde{0}$;

(b) Either:

(i) For all integers $k \in \mathcal{R}^+$, $d(k) = (y_1, y_2, \ldots, y_n)$ where each $y_j \in \mathcal{Z}^+$ and $\sum_{j=1}^{n} y_j = k$; or

(ii) There is an integer $m \geq 0$ such that for all positive integers $k \leq m$,

$d(k) = (y_1, y_2, \ldots, y_n)$ where each $y_j \in \mathcal{Z}^+$ and $\sum_{j=1}^{n} y_j = k$, and for all integers $k > m$, $d(k) = d(m)$;

(c) If $x \in \mathcal{R}^+$ is not an integer, then $d(x) = d([x]+1)$, where $[x]$ is the integer part of x.

Definition 3: *Let $\tilde{\mathcal{D}} = (\mathcal{A}, \Phi, \mathcal{L}, \mathcal{D})$ be a dialectical system, and let d_L be an associated dialogue, that is, a possibly-infinite time-ordered sequence of locutions from $\mathcal{L}$ instantiated by the topics in Φ, uttered by agents in $\mathcal{A}$ in accordance with the rules of game $\mathcal{D}$. Suppose the t-th element of d_L is the instantiated locution $L_j(t, P_i, \phi_k)$, for $t = 1, 2, \ldots$. The dialogue path d associated with d_L is the function $d : \mathcal{R}^+ \to \mathcal{R}^{n+}$ obtained by setting $d(0) = \tilde{0}$, setting each $d(t)$ equal to: $d(t-1) + (0, \ldots, 0, 1, 0, \ldots, 0)$, where the non-zero co-ordinate corresponds to that axis labeled (P_i, ϕ_k, L_j); and where $d(x) = d([x]+1)$, when $x \in [t-1, t)$. We say that d is the dialogue path which* corresponds to *or* matches *d_L.*

It is easy to see that such a d is a dialogue path. Thus, for each legal dialogue we have an associated continuous directed path through $\mathcal{R}^{n+}$, starting from the origin, made up of straight-line segments each parallel to an axis of the space, and each one unit in length. Such paths may represent infinite (condition b i of Definition 2) or terminating (condition b ii) dialogues. In the latter case, we say that the path d is *a terminating path with terminal time-point m* or that the path *terminates at m*, where m is the integer mentioned in condition (b) (ii) of Definition 2. In this case, we also say the dialogue path is of length m. If the former case, we say that the path d is *non-terminating* or *infinite*. It will be useful to distinguish generic dialogue paths (as defined above) from those which obey the rules of the dialogue game, $\mathcal{D}$.

Definition 4: *Let $d : \mathcal{R}^+ \to \mathcal{R}^{n+}$ be a dialogue path. Suppose $\mathcal{A}$ is a set of agents of size p, Φ a set of topics of discussion of size q, and $\mathcal{D}$ a dialogue game protocol with a set of locutions $\mathcal{L}$ of size l, with $n = pql$, such that there is a legal dialogue between the agents in $\mathcal{A}$ concerning the topics in Φ and conducted according to the rules of $\mathcal{D}$ using the set of locutions $\mathcal{L}$ whose dialogue path matches d. Then we say that d is* a dialogue path under *or* is legal under *the dialectical system $(\mathcal{A}, \Phi, \mathcal{L}, \mathcal{D})$.*

In general, not all dialogue paths will correspond to legal dialogues. This is because the various combination, termination and commitment rules create subsets of $\mathcal{R}^{n+}$ which a legal dialogue path either cannot enter or must traverse. Because the combination rules typically specify which instantiated locutions may, may not or must be uttered depending on the previous utterances in the dialogue and the identity of the agent speaking, these forbidden and obligatory regions will differ at each round in the dialogue. Thus, because the histories of two dialogues at any one time may be different, one dialogue path may traverse a region which is forbidden to another dialoue path; a dialogue path may even traverse a region which is forbidden to itself later or earlier in the same dialogue. In the sections below we will refer to the Forbidden Region and the Termination

Region for a dialogue path under particular dialectical system at a particular time point, with the obvious meanings.

3.3 Path and system equivalence

Because our focus is on specific types of agent interactions, such as negotiations or persuasions, we are concerned to see what external action commitments are made in the course of a dialogue, and what information is transferred between participants to achieve these commitments. We therefore require some measure of these, which we obtain by first examining the syntax of each locution-type.

Definition 5: *Suppose $\tilde{\mathcal{D}} = (\mathcal{A}, \Phi, \mathcal{L}, \mathcal{D})$ is a dialectical system, with $L \in \mathcal{L}$ a locution-type. We define the information possibly transferred by speakers uttering locution L, denoted* $\mathrm{Poss\_Info}(L)$*, by the set of all possible subsets of Φ which could be instantiated into L. Similarly, the actions possibly committed to by participants uttering locution L, denoted* $\mathrm{Poss\_Acts}(L)$*, is the set of all possible subsets of Φ_a which could be instantiated as actions in L. A procedural locution L has* $\mathrm{Poss\_Info}(L) = \mathrm{Poss\_Acts}(L) = \{\emptyset\}$*, i.e. the set containing only the empty set.*

For example, the locution with syntax $seek\_price(k, P_i, \phi)$, with k is positive integer (a time stamp), $P_i \in \mathcal{A}$ an agent, and $\phi \in \Phi$ a proposition, will have *Poss_Info(seek_price)* $= \{\{\theta\} \mid \theta \in \Phi\}$, i.e. all singleton subsets of Φ. By contrast, the locution with syntax $prefer(k, P_i, \phi, \psi)$ with k and P_i as before, and with $\phi, \psi \in \Phi$ propositions, will have *Poss_Info(prefer)* $= \{\{\theta, \gamma\} \mid \theta, \gamma \in \Phi, \theta \neq \gamma\}$, i.e. all subsets of two distinct elements of Φ. Thus, *Poss_Info(seek_price)* $\neq$ *Poss_Info(prefer)*. Thus, each locution defines a subset of $\mathcal{P}(\Phi) \times \mathcal{P}(\Phi_a)$. Each utterance of an information-transferring locution L in a given dialogue is instantiated with the contents of one element of the subset *Poss_Info(L)* of $\mathcal{P}(\Phi)$. Considering instantiation leads us to define the information actually transferred and actions actually committed to by participants in a given dialogue.

Definition 6: *Let d be a dialogue path under a dialectical system $\tilde{\mathcal{D}} = (\mathcal{A}, \Phi, \mathcal{L}, \mathcal{D})$, and let d_L be the corresponding dialogue. We define the* information transferred *by d, denoted* $\mathrm{Info}(d)$*, as the set of information transferred by the locutions uttered in d_L, i.e.*

$\mathrm{Info}(d) = \{\phi \in \Phi \mid \exists \mathcal{L}_j(t, P_i, \phi) \in d_L\}$.

Likewise, we define the actions committed to *by d, denoted* $\mathrm{Acts}(d)$*, as the set of action propositions committed to by the speakers of locutions uttered in d_L, i.e.*

$\mathrm{Acts}(d) = \{\phi \in \Phi_a \mid \exists \mathcal{L}_j(t, P_i, \phi) \in d_L\}$.

Using these definitions, we next define a notion of "closeness" of two dialogue paths.

Definition 7: *Let $d, e : \mathcal{R}^+ \to \mathcal{R}^{n+}$, be two legal dialogue paths under the same* *dialectical system* $(\mathcal{A}, \Phi, \mathcal{L}, \mathcal{D})$. *We say that* d is close to e *precisely in the case when both paths are terminating and Info(d) = Info(e) and Acts(d) = Acts(e).*

Note that two close paths may terminate at different time-points, i.e. be of different length. We have the following result, whose straightforward proof is omitted.

Proposition 1: *The relation of closeness between two legal dialogue paths is an equivalence relation.* □

We can therefore speak of two dialogue paths d and e under the same dialogue system being *equivalent*, which we denote by $d \sim e$. We also refer to the corresponding dialogues being equivalent. We denote the equivalence class of a dialogue path d by $[d]$, which we call a *path-equivalence*. We also have the following.

Proposition 2: *For any dialectical system* $\tilde{\mathcal{D}}$*, the set of path-equivalence classes of legal dialogue paths under* $\tilde{\mathcal{D}}$ *is finite.*
Proof. This result follows from the assumption that the set Φ is finite. □

Using the notion of path-equivalence, we now define a relationship of similarity between dialectical systems.

Definition 8: *Suppose* $\tilde{\mathcal{D}} = (\mathcal{A}_{\mathcal{D}}, \Phi_{\mathcal{D}}, \mathcal{L}_{\mathcal{D}}, \mathcal{D})$ *and* $\tilde{\mathcal{E}} = (\mathcal{A}_{\mathcal{E}}, \Phi_{\mathcal{E}}, \mathcal{L}_{\mathcal{E}}, \mathcal{E})$ *are two dialectical systems of dimension* m *and* n *respectively. We say that* $\tilde{\mathcal{D}}$ *is* similar *to* $\tilde{\mathcal{E}}$ *if there exists a one-to-one and onto function* $h : \mathcal{R}^{m+} \rightarrow \mathcal{R}^{n+}$ *such that for every terminating dialogue path* d *under* $\tilde{\mathcal{D}}$ *there is a terminating dialogue path* e *under* $\tilde{\mathcal{E}}$ *with:*

$h(d) = e$,
$\text{Info}(d) = \text{Info}(e)$,
$\text{Acts}(d) = \text{Acts}(e)$,

and such that for all dialogue paths $e' \sim e$ *under* $\tilde{\mathcal{E}}$*, there exists a dialogue path* $d' \sim d$ *such that* $h(d') = e'$.

In other words, two dialogue systems are similar if the first can be mapped to the second so that terminating dialogues are mapped to terminating dialogues while preserving information-transfers and action-commitments, and so that equivalent dialogue paths are mapped to equivalent dialogue paths. Clearly, for such a map h to exist, the two universes of discourse Φ_D and Φ_E must intersect, as must their respective action subsets. The second condition in the definition of h, the existence of a d' for every e' with $h(d') = e'$, may be considered as analogous to a continuity requirement on h, since dialogue paths which are close to one another in the second dialectical system are required to be the images of dialogue paths close to one another in the first. As with path-equivalence, our notion of similarity of dialectical systems is an equivalence relationship, a statement whose straightforward proof we omit.

Proposition 3: *The relationship $\sim$ between dialectical systems is an equivalence relation.* $\square$

We can therefore speak of two dialectical systems $\tilde{\mathcal{D}}$ and $\tilde{\mathcal{E}}$ being *equivalent*, denoted $\tilde{\mathcal{D}} \sim \tilde{\mathcal{E}}$. This notion of equivalence of dialectical systems is a global property defined in terms of the existence of local properties, i.e. similarity of dialogue paths. Note that the equivalence mapping h maps legal dialogue paths to legal dialogue paths; since such paths avoid Forbidden Regions in their respective spaces, then h preserves this structure in mapping $\mathcal{R}^{m+}$ to $\mathcal{R}^{n+}$.

4 Comparing Dialectical Systems

In this section, we begin with a connection between the equivalence of dialectical systems and the sets of equivalence classes of the dialogue paths under them.

Proposition 4: *Suppose $\tilde{\mathcal{D}} \sim \tilde{\mathcal{E}}$ are two equivalent dialectical systems. Then the respective sets of path-equivalence classes generated by legal terminating dialogue paths under each system are isomorphic.*
Proof. We need to show that there is a one-to-one and onto map between the two sets of path-equivalence classes. Let $\Phi = \Phi_D \cup \Phi_E$ be the union of the two universes of discourse, and $\Phi_a = \Phi_D a \cup \Phi_E a$ the union of the two subsets of action propositions. By the definition of path-equivalence, each class in the set of path-equivalence classes for a specific dialectical system corresponds to a unique subset of $\mathcal{P}(\Phi) \times \mathcal{P}(\Phi_a)$, where $\mathcal{P}(X)$ is the power set of X. For each path-equivalence under $\tilde{\mathcal{D}}$, assign to it the path-equivalence under $\tilde{\mathcal{E}}$ corresponding to the same subset of $\mathcal{P}(\Phi) \times \mathcal{P}(\Phi_a)$. Such a path-equivalence class under $\tilde{\mathcal{E}}$ exists because the equivalence mapping h from $\tilde{\mathcal{D}}$ to $\tilde{\mathcal{E}}$ preserves the information transmitted and the actions committed to by each dialogue. This mapping is one-to-one because each path-equivalence class is associated with a unique subset of $\mathcal{P}(\Phi) \times \mathcal{P}(\Phi_a)$.

To prove it is onto, suppose, for purposes of contradiction, there is some path-equivalence class $[e]$ under $\tilde{\mathcal{E}}$ which corresponds to a subset of $\mathcal{P}(\Phi) \times \mathcal{P}(\Phi_a)$ to which no path-equivalence class under $\tilde{\mathcal{D}}$ corresponds. Consider a dialogue path e in the path-equivalence class $[e]$. Because $\tilde{\mathcal{D}} \sim \tilde{\mathcal{E}}$, then there must be a dialogue path d under $\tilde{\mathcal{D}}$ such that $h(d) = e$. Thus, *Info(d)* $=$ *Info(e)* and *Acts(d)* $=$ *Acts(e)*. But this just means that $[d]$ is associated with the same subset of $\mathcal{P}(\Phi) \times \mathcal{P}(\Phi_a)$ as is $[e]$, thus contradicting the assumption. $\square$

Proposition 4 shows that dialogue equivalence, which is defined in terms of mappings between real spaces, preserves the structure of the sets of associated path-equivalence classes. This is really not surprising given the definition of dialogue equivalence. However, it allows us to deduce the following interesting corollary:

Proposition 5: *Suppose* $\tilde{\mathcal{D}} = (\mathcal{A}, \Phi, \mathcal{L}_{\mathcal{D}}, \mathcal{D})$ *and* $\tilde{\mathcal{E}} = (\mathcal{A}, \Phi, \mathcal{L}_{\mathcal{E}}, \mathcal{E})$ *are two dialectical systems of dimension* m *and* n *respectively, such that* $\mathcal{L}_{\mathcal{E}} = \mathcal{L}_{\mathcal{D}} \cup \{L'\}$. *Moreover, suppose that* $\{\emptyset\} \neq \text{Poss\_Info}(L') \neq \bigcup_{L_j \in J} \text{Poss\_Info}(L_j)$, *for all* $J \subseteq \mathcal{L}_{\mathcal{D}}$. *Then* $\tilde{\mathcal{D}}$ *and* $\tilde{\mathcal{E}}$ *are not equivalent.*

Proof. If we had $\tilde{\mathcal{D}} \sim \tilde{\mathcal{E}}$ then, by Proposition 4, we would have an isomorphism of the two sets of path-equivalence classes. However, since $\{\emptyset\} \neq$ *Poss_Info*(L') $\neq \bigcup_{L_j \in J}$ *Poss_Info*(L_j), for all $J \subseteq \mathcal{L}_{\mathcal{D}}$, there is at least one instantiation of locution L' which transfers a non-empty subset of $\mathcal{P}(\Phi)$ not transferred by any dialogue without this locution. Since the dialogues of system $\tilde{\mathcal{D}}$ consist only of the locutions in $\mathcal{L}_{\mathcal{D}}$, we thus have a contradiction. Hence $\tilde{\mathcal{D}}$ and $\tilde{\mathcal{E}}$ are not equivalent. □

A similar result applies to dialogue systems which differ only by a locution which commits to actions not committed to by the other locutions. Moreover, both results apply if there is more than one locution $L' \in \mathcal{L}_{\mathcal{E}} \setminus \mathcal{L}_{\mathcal{D}}$ which transfers information or commits to actions not possible using the locutions in $\mathcal{L}_{\mathcal{D}}$. Similarly, we have the following corollary, whose similar proof we omit.

Proposition 6: *Suppose* $\tilde{\mathcal{D}} = (\mathcal{A}, \Phi, \mathcal{L}_{\mathcal{D}}, \mathcal{D})$ *and* $\tilde{\mathcal{E}} = (\mathcal{A}, \Phi, \mathcal{L}_{\mathcal{E}}, \mathcal{E})$ *are two dialectical systems of dimension* m *and* n *respectively, such that the protocols* $\mathcal{D}$ *and* $\mathcal{E}$ *differ only by a rule in* $\mathcal{E}$ *which terminates a dialogue under some conditions. Then* $\tilde{\mathcal{D}} \sim \tilde{\mathcal{E}}$ *only if for each terminating dialogue-path* e *under* $\mathcal{E}$ *in which the rule is invoked and leads to termination there is a terminating dialogue path* d *under* $\mathcal{D}$ *with* $\text{Info}(d) = \text{Info}(e)$ *and* $\text{Acts}(d) = \text{Acts}(e)$. □

These corollaries are important because they provide us with some guidance for the design of dialogue systems. With them, we know that adding a locution or a termination rule which transfers information not transferred in the current set of locutions will create a non-equivalent system, i.e. a dialogue system in which there will be dialogues transferring different information or committing to different actions. Dialogue-system equivalence, as we have defined it here, is not the only criterion one could use for design and assessment of dialogue game protocols; one may wish to add or not add such locutions or rules for other reasons, such as overall simplicity or to encode a particular theory of argumentation [21]. However, we believe dialogue equivalence is an important criterion and these results give us a purchase on understanding its implications for dialogue-game protocol design.

5 Finite and Infinite Dialogues

Dialogues conducted under a dialogue protocol may terminate after some finite number of locutions, or they may, if the protocol rules so allow, continue indefinitely. Since most software agents are intended for real-world applications, it is unlikely that agents will be permitted by their human principals to participate in a dialogue indefinitely. Every agent will eventually have to withdraw from a non-terminating dialogue, no matter how desirable the agent or its principal believes achievement of the dialogue's objectives to be. But when should a withdrawal occur? To explore this question, we consider the relationship between finite (terminating) and infinite dialogues under a protocol. We do this from the perspective of the Ehrenfeucht-Fraïssé games of mathematical logic [13]. These consider the question of the isomorphism of two infinite structures by assessing whether isomorphisms exist between partial, finite subsets of the structures.[7] We start by defining the sub-sets of dialogues under a protocol which terminate within a given number of utterances.

Definition 9: *Let $\tilde{\mathcal{D}}$ be a dialectical system and let n be a positive integer. Then we define $\tilde{\mathcal{D}}_n$ to be the set of all legal dialogues conducted under $\tilde{\mathcal{D}}$ which terminate by time t, for $0 \leq t \leq n$. Further, in an abuse of notation, let $[\tilde{\mathcal{D}}_n]$ denote the set of path-equivalence classes of the dialogues contained in $\tilde{\mathcal{D}}_n$, and let $[\tilde{\mathcal{D}}]$ denote the set of path-equivalence classes of the dialogues contained in $\tilde{\mathcal{D}}$. We call $[\tilde{\mathcal{D}}_n]$ a finite path-equivalence class set and $[\tilde{\mathcal{D}}]$ the path-equivalence class set for $\tilde{\mathcal{D}}$.*

Then we have the following proposition connecting finite dialogues to all possible dialogues under a protocol.

Proposition 7: *Suppose $\tilde{\mathcal{D}}$ is a dialectical system. Then there exists m, a non-negative integer, such that for all integers $n > m$, $[\tilde{\mathcal{D}}_n]$ is isomorphic to $[\tilde{\mathcal{D}}]$.*
Proof. Each element of the set $[\tilde{\mathcal{D}}]$ is the equivalence class of some finite dialogue under $\tilde{\mathcal{D}}$; such a dialogue must eventually terminate and thus be included in $[\tilde{\mathcal{D}}_n]$, for some n. But the set $[\tilde{\mathcal{D}}]$ is finite, by Proposition 2. The result follows. □

As a corollary to this proposition, we can show that equivalent dialectical systems eventually have isomorphic finite path-equivalence class sets.

Proposition 8: *Suppose $\tilde{\mathcal{D}} \sim \tilde{\mathcal{E}}$ are two equivalent dialectical systems. Then there exists m, a non-negative integer, such that for all integers $n > m$, $[\tilde{\mathcal{D}}_n]$ is isomorphic to $[\tilde{\mathcal{E}}_n]$.*
Proof. We prove this by contradiction. Assuming the result is not true means that for every positive integer m, there exists a larger integer n such that $[\tilde{\mathcal{D}}_n]$

[7]The concept of partial, finite isomorphisms between two infinite structures was due to Roland Fraïssé [10]. Andrzej Ehrenfeucht [7] extended this idea by constructing abstract games of $2n$ moves between two players, *Duplicator* and *Spoiler*, who alternate in choosing elements of the two infinite structures for a finite number of moves. The objective of Duplicator is to show that the two structures are isomorphic, while Spoiler has the objective of showing they are not. Duplicator wins a game of $2n$ moves if the two subsets of n elements selected by the players from the two parent structures are isomorphic. Spoiler wins otherwise. If the two infinite structures are in fact isomorphic, then a winning strategy exists for Duplicator for the game of size $2n$, for every finite n. However, the converse need not be true, as one could readily imagine isomorphisms existing between every pair of equal-sized finite subsets without there being an (infinite) isomorphism between the parent structures.

is not isomorphic to $[\tilde{\mathcal{E}}_n]$. But the two dialectical systems are equivalent, and so their respective (infinite) path-equivalence class sets are isomorphic (Proposition 4). But both these sets are finite (Prop. 2). Thus, there must eventually be some integer value m after which neither of the respective sequences of finite path-equivalence sets increases in size. Thus, both sequences must converge on the infinite sets after a finite time, which contradicts the assumption that not all such pairs are isomorphic. □

Recall that the path-equivalence classes are defined in terms of information transmitted or action commitments incurred in the course of a dialogue. Thus, if an agent's objectives are defined in such terms — receipt or transfer of information, or incurrence of an action commitment — then Proposition 7 says that every possible dialogue objective is achievable in some dialogue or other in a finite time. Of course, although consoling, this may provide no guidance to an agent now engaged in a particular dialogue about whether to continue or to leave; we hope to return to this question in future, perhaps linking it to recent research in AI on the reconsideration of agent intentions [34].

6 Conclusion

In this paper we have defined a novel geometric semantics for dialogue game protocols for interactions between autonomous agents and begun to explore its formal properties. We have used these properties to show that two protocols which differ only in one locution are not equivalent if that additional locution transmits information or commits its speaker to actions which are not expressible by any combination of the other locutions. We have also begun to explore relationships between finite and infinite dialogues. Although our methods are not sophisticated mathematically, we believe our results are important because of the formal guidance they provide for the design and evaluation of dialogue game protocols. We believe this application domain contains further potential for investigation. In addition to the questions raised above, there are two other issues on which we are focused. Firstly, in the work presented here we have not used the fact that dialogue paths are time-directed. The recent theory of *directed homotopies*, or *dihomotopies* [8, 27], which has arisen in applying algebraic topology to problems of concurrency in theoretical computer science, may be relevant as we consider the time-direction of dialogues. Secondly, we believe the results of this paper could be readily expressed in category-theoretic terms; indeed, category theory may provide a more elegant means to abstract away from non-essential features of a protocol, and to reason about classes of protocols. We are pursuing both these lines of inquiry.

Acknowledgments

This reseaerch was partly funded by the EU IST Programme through the Sustainable Lifecycles in Information Ecosystems (SLIE) Project (IST-1999-10948) and by the British Engineering and Physical Sciences Research Council (EP-SRC) through a PhD studentship. We are also grateful for comments from Mark Johnson and Grant Malcolm, the audience at the GETCO-2001 workshop in Aalborg and the anonymous referees.

References

[1] S. Abramsky. Semantics of interaction: an introduction to game semantics. In A. M. Pitts and P. Dybjer, editors, *Semantics and Logics of Computation*, pages 1–31. Cambridge University Press, Cambridge, UK, 1997.

[2] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In E. Durfee, editor, *Proceedings of the Fourth International Conference on Multi-Agent Systems (ICMAS 2000)*, pages 31–38, Boston, MA, USA, 2000. IEEE Press.

[3] Aristotle. *Topics.* Clarendon Press, Oxford, UK, 1928.

[4] F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Agent theory for team formation by dialogue. In C. Castelfranchi and Y. Lespérance, editors, *Pre-Proceedings of the Seventh International Workshop on Agent Theories, Architectures, and Languages (ATAL-2000)*, pages 141–156, Boston, USA, 2000.

[5] F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Creating collective intention through dialogue. *Logic Journal of the IGPL*, 9(2):305–319, 2001.

[6] F. Dürrenmatt. *A Dangerous Game.* Jonathan Cape, London, UK, 1960. (Translation by R. and C. Winston of *Die Panne*, published in German in 1956.).

[7] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Mathematicae*, 49:129–141, 1961.

[8] L. Fajstrup. Loops, ditopology and deadlocks. *Mathematical Structures in Computer Science*, 10:459–480, 2000.

[9] FIPA. Communicative Act Library Specification. Technical Report XC00037H, Foundation for Intelligent Physical Agents, 10 August 2001.

[10] R. Fraïssé. Sur les classifications des systems de relations. *Publications Scientifiques de l'Universite d'Alger*, 1(1), juin 1954.

[11] C. L. Hamblin. *Fallacies.* Methuen, London, UK, 1970.

[12] D. Hitchcock, P. McBurney, and S. Parsons. A framework for deliberation dialogues. In H. V. Hansen, C. W. Tindale, J. A. Blair, and R. H. Johnson, editors, *Proceedings of the Fourth Biennial Conference of the Ontario Society for the Study of Argumentation (OSSA 2001)*, Windsor, Ontario, Canada, 2001.

[13] W. Hodges. *A Shorter Model Theory.* Cambridge University Press, Cambridge, UK, 1997.

[14] J. Hulstijn. *Dialogue Models for Inquiry and Transaction.* PhD thesis, Universiteit Twente, Enschede, The Netherlands, 2000.

[15] N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, M. Wooldridge, and C. Sierra. Automated negotiation: prospects, methods and challenges. *International Journal of Group Decision and Negotiation*, 10(2):199–215, 2001.

[16] J. A. Levin and J. A. Moore. Dialogue-games: metacommunications structures for natural language interaction. *Cognitive Science*, 1(4):395–420, 1978.

[17] P. Lorenzen and K. Lorenz. *Dialogische Logik.* Wissenschaftliche Buchgesellschaft, Darmstadt, Germany, 1978.

[18] J. D. MacKenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133, 1979.

[19] J. Mayfield, Y. Labrou, and T. Finin. Evaluating KQML as an agent communication language. In M. J. Wooldridge, J. P. Müller, and M. Tambe, editors, *Intelligent Agents II*, volume 1039 of *Lecture Notes in Artificial Intelligence*, pages 347–360, Berlin, Germany, 1996. Springer.

[20] P. McBurney and S. Parsons. Agent ludens: games for agent dialogues. In S. Parsons and P. Gmytrasiewicz, editors, *Game-Theoretic and Decision-Theoretic Agents: Proceedings of the 2001 AAAI Spring Symposium*, pages 70–77, Menlo Park, CA, USA, 2001. AAAI Press. Technical Report SS-01-03.

[21] P. McBurney and S. Parsons. Desiderata for dialogue-game protocols for agent interactions. In S. Parsons and A. Hunter, editors, *Adventures in Argumentation: Proceedings of a Workshop held with the Sixth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU-2001)*, pages 44–47, Toulouse, France, 2001. Institut de Recherche en Informatique de Toulouse (IRIT).

[22] P. McBurney and S. Parsons. Representing epistemic uncertainty by means of dialectical argumentation. *Annals of Mathematics and Artificial Intelligence*, 32(1–4):125–169, 2001.

[23] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 2002 (*In press*).

[24] P. McBurney, R. van Eijk, S. Parsons, and L. Amgoud. A dialogue-game protocol for agent purchase negotiations. *Journal of Autonomous Agents and Multi-Agent Systems*, 2002 (*In press*).

[25] P. Mittelstaedt. *Quantum Logic*. D. Reidel, Dordrecht, The Netherlands, 1979.

[26] S. Parsons, C. Sierra, and N. R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.

[27] M. Raussen. On the classification of dipaths in geometric models for concurrency. *Mathematical Structures in Computer Science*, 10(4):427–457, 2000.

[28] F. Sadri, F. Toni, and P. Torroni. Logic agents, dialogues and negotiation: an abductive approach. In M. Schroeder and K. Stathis, editors, *Proceedings of the Symposium on Information Agents for E-Commerce, Artificial Intelligence and the Simulation of Behaviour Conference (AISB-2001)*, York, UK, 2001. AISB.

[29] K. Sycara. Persuasive argumentation in negotiation. *Theory and Decision*, 28:203–242, 1990.

[30] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Press, Albany, NY, USA, 1995.

[31] H. Wohlrapp. A new light on non-deductive argumentation schemes. *Argumentation*, 12:341–350, 1998.

[32] M. J. Wooldridge. *Reasoning about Rational Agents*. MIT Press, Cambridge, MA, USA, 2000.

[33] M. J. Wooldridge. Semantic issues in the verification of agent communication languages. *Journal of Autonomous Agents and Multi-Agent Systems*, 3(1):9–31, 2000.

[34] M. J. Wooldridge and S. Parsons. Intention reconsideration reconsidered. In J. P. Müller, M. Singh, and A. Rao, editors, *Intelligent Agents V*, Lecture Notes in Artificial Intelligence 1365, pages 63–79, Berlin, Germany, 1999. Springer.