When are two protocols the same?

Mark W. Johnson[1], Peter McBurney[2], and Simon Parsons[23]

[1] Department of Mathematics
Pennsylvania State University Altoona
Altoona PA 16601-3760 USA
mwj3@psu.edu
[2] Department of Computer Science
University of Liverpool
Liverpool L69 7ZF UK
{p.j.mcburney,s.d.parsons}@csc.liv.ac.uk
[3] Department of Computer and Information Science
Brooklyn College
City University of New York
Brooklyn NY 11210 USA
parsons@sci.brooklyn.cuny.edu

Abstract. A number of protocols based on the formal dialogue games of philosophy have recently been proposed for interactions between autonomous agents. Several of these proposals purport to assist agents engaged in the same types of interactions, such as persuasions and negotiations, and are superficially different. How are we to determine whether or not these proposals are substantially different? This chapter considers this question and explores several alternative definitions of equivalence of protocols.

1 Introduction

Recently, several authors have proposed agent communications protocols based on the formal dialogue games of philosophy. These are interactions between two or more players, where each player "moves" by making utterances, selected from a finite set of possible locutions, according to a defined set of rules. These games have been studied by philosophers since at least the time of Aristotle [5], and have recently been the focus of renewed attention in philosophy [15].

Because dialogues conducted according to these games are rule-governed, they have been of particular interest to computer scientists, for modeling human-computer interaction [7], for modeling complex reasoning, such as that in legal domains [24], and for the task of software specification [10]. Recently, dialogue game formalisms have been proposed as protocols for interaction between autonomous software agents. In this domain, such protocols have been proposed for agent dialogues involving: persuasion [1, 8, 9]; joint inquiry [18]; negotiations over the division of scarce resources [4, 22, 30]; deliberations over a course of action [16]; and discovery of rare events [17].

As can be seen from this list of citations, more than one dialogue game protocol has sometimes been proposed for the same type of agent interaction. This presents a user or potential user of such protocols with a number of questions, including:

– How might one choose between two protocols?
– When is one protocol preferable to another?
– When do two protocols differ?

The first of these questions is of interest to agent designers who are considering how to allow their agents to interact—how do they pick one protocol from the many that have been proposed? Answering it involves, at the very least, having some way of describing the various features of protocols. In other work, two of us have taken a step towards answering this question by proposing a set of desirable properties for dialogue game protocols for agent interactions [21], and assessing various protocols against these properties.[1]

Such assessments could also provide a partial answer to the second question above. This question is again of interest to agent designers for much the same reason as the first, but goes a little further in that it also requires an understanding of what makes protocols good, and what makes them good for particular tasks (since it seems clear from the wide variety of extant protocols that some are good for some tasks and others are good for other tasks).

However, perhaps a more fundamental question is the third question given above—when do two protocols differ. Not only is an answer to this a prerequisite to being able to choose between two protocols, but it is also essential if we are to be able to tell if a protocol is new (in the sense of providing a different functionality from an existing protocol rather than just having equivalent locutions with different names) and if a protocol conforms to some specification, such as that laid down as the standard for interacting within some electronic institution [29]. Mention of functionality of a protocol leads naturally to considerations of semantics: what effects does a protocol, or rather, the dialogues conducted under it, have?[2]

In this paper, we present a preliminary exploration of this third question, including several alternative definitions for protocol equivalence. In order to do this, we first need to define what we mean by a dialogue game protocol (defined in Section 2) and to classify types of dialogues and types of locutions. We then present a classification of dialogue types, essentially a standard one from the philosophy of argumentation, due to Walton and Krabbe. Our classification of agent dialogue locutions (presented in Section 3), although based on a typology of speech acts due to Habermas, is novel. Section 4 presents our various definitions of dialogue game protocol equivalence, and explores their relationships to one another. Section 5 discusses possible extensions of the work of this paper.

2 Dialogue Game Protocols

Elsewhere [19], we identified the key elements of a dialogue game protocol, in a generic model of a formal dialogue game. We assume the dialogue occurs between autonomous

[1] For comparison, we also assessed the FIPA ACL [11] against these criteria.

[2] Thus our focus differs from that of traditional communications theory, which ignores the semantics of messages, e.g., *"Frequently the messages have* meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem."* [32, p. 31].

software agents, and that the topics of their discussion can be represented in some logical language; we represent these topics with well-formed formulae denoted by lower-case Greek letters, θ, ϕ, etc. The specification of a dialogue game protocol then requires specification of:

Commencement Rules: Rules that define the circumstances under which the dialogue commences.

Locutions: Rules that indicate what utterances are permitted. Typically, legal locutions permit participants to assert propositions, permit others to question or contest prior assertions, and permit 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. The dialogue game rules may also permit participants to utter propositions to which they assign differing degrees of belief or commitment, for example: one may merely *propose* a proposition, a speech act which entails less commitment than would an *assertion* of the same proposition.

Combination Rules: Rules that 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 θ and subsequently the proposition $\neg\theta$ in the same dialogue, without in the interim having retracted the former assertion.

Commitments: Rules that define the circumstances under which participants express commitment to a proposition in the dialogue. Typically, the assertion of a claim θ in the debate is defined as indicating to the other participants some level of commitment to, or support for, the claim. Since [15], formal dialogue systems typically establish and maintain public sets of commitments, called *commitment stores*, for each participant. When a participant utters a locution which incurs a commitment, the corresponding proposition is inserted into that participant's commitment store, where it is visible to the other participants. These stores are usually non-monotonic, in the sense that participants can also retract committed claims, although possibly only under defined circumstances. Fully specifying a protocol includes specification of the ways in which potentially-conflicting commitments may interact with one another.

Termination Rules: Rules that define the circumstances under which the dialogue ends.

Some comments on this model are in order. Firstly, the circumstances which may lead to the commencement of a specific dialogue under a given protocol are, strictly speaking, not part of that dialogue or protocol. Accordingly, it is reasonable to consider a meta-dialogue, where discussions about which dialogues to enter are undertaken [19] or a hierarchy of nested sequential dialogues [27]. Alternatively, participating agents may select a dialogue from some agreed library of dialogue types, as in [6]. Secondly, dialogue games are different from conversation policies [12], which are short sequences of legal utterances with a common purpose. Thus a conversation policy sits between a single utterance and a complete dialogue in length; it governs a portion of a complete dialogue, rather than the whole, as is the case with a dialogue game.

Thirdly, it is worth noting that more than one notion of *commitment* is present in the literature on dialogue games. For example, Hamblin treats commitments in a purely dialogical sense: *"A speaker who is obliged to maintain consistency needs to keep a store of statements representing his previous commitments, and require of each new statement he makes that it may be added without inconsistency to this store. The store represents a kind of persona of beliefs; it need not correspond with his real beliefs . . ."* [15, p. 257]. In contrast, Walton and Krabbe [36, Chapter 1] treat commitments as obligations to (execute, incur or maintain) a course of action, which they term action commitments. These actions may be utterances in a dialogue, as when a speaker is forced to defend a proposition he has asserted against attack from others; so Walton and Krabbe also consider propositional commitment as a special case of action commitment [36, p. 23]. As with Hamblin's treatment, such dialogical commitments to propositions may not necessarily represent a participant's true beliefs. In contrast, Singh's social semantics [35], requires agent participants to an interaction to express publicly their beliefs and intentions, and these expressions are called *social commitments*. These include both expressions of belief in some propositions and expressions of intent to execute or incur some future actions.[3]

Our primary motivation is the use of dialogue games as the basis for interaction protocols between autonomous agents. Because such agents will typically enter into these interactions in order to achieve some wider objectives, and not just for the enjoyment of the interaction itself, we believe it is reasonable to define commitments in terms of future actions or propositions external to the dialogue. In a commercial negotiation dialogue, for instance, the utterance of an offer may express a willingness by the speaker to undertake a subsequent transaction on the terms contained in the offer. For this reason, we can view commitments as mappings between locutions and subsets of some set of action-statements.

Classifying Dialogues

What different sorts of agent dialogues are there? If we assume that agents enter dialogues with each other in order to achieve specific objectives, it would seem reasonable to classify the dialogues in terms of the private and shared objectives of the participants. Indeed, these criteria — the private objectives and the shared objectives — were used by argumentation theorists Doug Walton and Erik Krabbe in their influential typology of human dialogues [36].[4] In addition, their typology is based on what information each participant has at the commencement of the dialogue (of relevance to the topic under discussion). The result was six primary types of dialogue, as follows: **Information-Seeking Dialogues** are those where one participant seeks the answer to some question(s) from another participant, who is believed by the first to know the answer(s). In

[3] It is worth noting that all these notions of *commitment* differ from that commonly used in discussion of agent's internal states, namely the idea of the persistence of a belief or an intention [38, p. 205]. As Singh [33] argues, there is a qualitative difference between social commitments of the kind discussed here, and personal commitments of the kind encoded in beliefs, desires, and intentions. He further argues that one kind of commitment cannot be derived from another.

[4] This typology is described in more detail in another chapter of this volume [20].

Inquiry Dialogues the participants collaborate to answer some question or questions whose answers are not known to any one participant. **Persuasion Dialogues** involve one participant seeking to persuade another to accept a proposition he or she does not currently endorse. In **Negotiation Dialogues**, the participants bargain over the division of some scarce resource. Here, the goal of the dialogue — a division of the resource acceptable to all — may be in conflict with the individual goals of the participants – to maximize their individual shares. Participants of **Deliberation Dialogues** collaborate to decide what action or course of action should be adopted in some situation. Here, participants share a responsibility to decide the course of action, or, at least, they share a willingness to discuss whether they have such a shared responsibility. Note that the best course of action for a group may conflict with the preferences or intentions of each individual member of the group; moreover, no one participant may have all the information required to decide what is best for the group. In **Eristic Dialogues**, participants quarrel verbally as a substitute for physical fighting, aiming to vent perceived grievances.

For our purposes, we note that the termination rules for these different dialogue types can be expressed succinctly in terms of utterances of support within each dialogue for certain propositions. An Information-seeking dialogue, for example, can terminate normally once the participant who sought the answer to some question indicates publicly that a given proposition provides a satisfactory answer. For normal termination of an Inquiry dialogue, all (or some designated subset) of participants must express such public indication. For a Persuasion dialogue, normal termination will occur when the participant being persuaded publicly endorses (via an appropriate locution) support for the proposition at issue. Similarly, normal termination rules for Negotiation and Deliberation dialogues may be be articulated in terms of participant support for particular propositions — in these two cases, for propositions which express commitments to future actions external to the dialogues.

3 Classifying Locutions

The agent dialogue types above involve agents seeking to reach a common understanding of some situation or a collective agreement to undertake certain actions. The theory of Communicative Action of the philosopher Jürgen Habermas [14] attempts to understand how human participants achieve these goals through dialogue, and, as part of this theory, Habermas proposes a typology of statements made in such dialogues [14, pp. 325–326].[5] We have used this typology as the basis for our own classification of locutions in agent dialogues (with Habermas' labels given in parentheses):

Factual Statements (Constative Speech Acts): These are statements which seek to represent the state of the objective, external world. Statements of belief about factual matters are examples of these utterances. Contesting such a statement means denying that it is a true description of the reality external to the dialogue.

Value Statements (Expressive Acts): These are statements which seek to represent the state of the speaker's internal world, i.e., they reveal publicly a subjective preference or value assignments. Such statements may only be contested by doubting the sincerity of the speaker.

[5] Habermas' classification is derived from the typology of speech acts of Searle [31].

Connection Statements (Regulative Acts): These are statements which assert some social or other relationship between different participants.

Requests: These are statements about a desired state of the external world, in which an agent seeks another agent to act so as to bring about this state. Requests may be criticized on the grounds of effectiveness, i.e., that the requested action will not, in fact, bring about the desired world state. In addition, they may be refused, with or without a reason being expressed.

Promises: These are statements about a desired state of the external world, in which an agent itself agrees to act so as to bring about this state. As with requests, promises may be criticized on the grounds of effectiveness, and may be accepted or rejected, with or without reasons.

Inferences (Operative Acts): These are statements which refer to the content of earlier statements in a dialogue, drawing inferences from them or assessing their implications. Contestation of such statements can take the form of questioning the appropriateness or the validity of the inferences made.[6]

Procedural Statements (Communicative Acts): These are statements about the activity of dialogue itself, such as the rules for participation and debate. In many human discourses, these often themselves become the focus of debate, dominating the issues of substance. In some dialogues, the participants may agree to submit such issues to a chairperson or other authority for determination.

By distinguishing between Requests and Promises in the way we have, our classification differs from that of Habermas. He does not include promises in his structure, and requests are treated as commands (Imperative Acts) rather than as requests. We believe our approach is more appropriate in a context of agent autonomy. Complete autonomy, as may occur for example in open multi-agent systems, means that imperative statements may have no force: other agents can only be requested to perform some action, and never commanded to do it. In closed multi-agent systems, agents may have an hierarchical relationship with one another, and so not have complete autonomy, as for example, when they represent different departments of the same company. However, even in such applications, agents may still exercise some autonomy over a limited domain, and so a classification which includes both Requests and Promises is appropriate.

Of the seven types of locutions in this typology, only Factual statements, Promises and Requests relate to propositions with referants in the world external to the dialogue: Factual statements express beliefs about the world, while Promises and Requests concern propositions linked to actions in the world. Of these, only promises seek to *change* the external world in any way. Promises may be made by more than one participant (i.e., they may be joint promises) and may be viewed as commitments to create or to maintain a state of the world in which specified propositions are true [34, 35]. In the next section, we will map Factual Statements and Promises to sets of propositions which represent external reality.

[6] Our definition departs slightly from that of Habermas, in that we permit Inferential Statements to have *"genuine communicative intent."*

4 Protocol Similarity

4.1 Concepts of Equivalence

In this section, we explore the question as to when two dialogue game protocols may be considered the same. To fix ideas, we first assume a finite set $\mathcal{A} = \{P_i | i = 1, \ldots, p\}$, of dialogue participants, or agents. Dialogues conducted by this set of agents are assumed to concern a finite set $\Phi = \{\phi_i | i = 1, \ldots, q\}$ of well-formed formulae in some propositional language, which we call the set of discussion topics. For this paper, both the set of agents and the set of topics are assumed fixed throughout. We denote dialogue game protocols by upper case script Roman letters, $\mathcal{D}, \mathcal{E}$, etc. Each protocol $\mathcal{D}$ comprises a finite set of legal locution-types, denoted $\mathcal{L}_{\mathcal{D}} = \{L_j | \ j = 1, \ldots, l\}$, and a number of combination, commitment and termination rules, denoted $\mathcal{R}_{\mathcal{D}} = \{R_j | \ j = 1, \ldots, r\}$. We assume that time can be represented by the non-negative real numbers, $[0, \infty)$, with locutions in a dialogue uttered simultaneously with the positive integers, i.e., the first utterance in the dialogue occurs at time $t = 1$, the second at time $t = 2$, etc. We define dialogues and partial dialogues as follows:

Definition 1. *A* dialogue $\tilde{d}$ *under dialogue-game protocol $\mathcal{D}$ is an ordered and possibly-infinite sequence of valid locutions, each possibly instantiated by one or more discussion topics, thus:*

$$\tilde{d} = (\ L_{d,t}(\theta_t) \mid t = 1, 2, \ldots)$$

with each $L_{d,t} \in \mathcal{L}_{\mathcal{D}}$ and each $\theta_t \in \Phi$. For any integer time-point $k > 0$, we say a partial dialogue to time k, $\tilde{d}_k$, *is an ordered and finite sequence of valid possibly-instantiated locutions $(\ L_{d,t}(\theta_t) \mid t = 1, \ldots, k)$, with each utterance valid under the combination rules of protocol $\mathcal{D}$, and such that $L_{d,k}(\theta_k)$ does not terminate the dialogue $\tilde{d}$ under the termination rules of $\mathcal{D}$.*

Drawing on the general structure of a dialogue game protocol presented in Section 2 we can make an initial attempt at defining protocol similarity as follows:

Definition 2. *(**Syntactic Equivalence**) Two protocols $\mathcal{D}$ and $\mathcal{E}$ are* syntactically equivalent *if their locutions, combination rules, commitment rules and termination rules are (respectively) the same, i.e., if $\mathcal{L}_{\mathcal{D}} = \mathcal{L}_{\mathcal{E}}$ and $\mathcal{R}_{\mathcal{D}} = \mathcal{R}_{\mathcal{E}}$.*

Thus, under this definition, two protocols are the same if their syntax is identical. This definition seems too strict, as it precludes us identifying two protocols which may differ in small but superficial ways, for example if one protocol has redundant locutions or rules.

Indeed, given a strictly syntactic notion of equality, it will classify two protocols which have sets of locutions which differ only in the names given to the locutions, as different. As an example of such a pair of dialogues, consider the two in [2] and [3]. The latter paper is a French language version of the former, and the protocol discussed in the two papers has locutions with exactly the same properties, but with different names (the names in the latter paper are the translation of the names in the former paper). The two protocols are not syntactically equivalent, despite the fact that they have exactly the same properties. Thus we need a less strict notion of equivalence. However, to achieve this we will need some notion of semantics, or meaning, for the dialogues under a protocol. We now present such a notion.

As mentioned in the previous section, Factual statements and Promises relate to propositions with referants in the world external to the dialogue. In each case, we may view the instantiated locution as invoking a subset of the elements of Φ, the set of discussion topics, and so each utterance comprising a Factual statement or a Promise defines a subset of Φ. For a given instantiated locution $L_j(\theta)$ in a protocol $\mathcal{D}$, we denote this subset by $\Phi_{\mathcal{D}}(L_j(\theta))$, and call it *the commitment set of* $L_j(\theta)$ *in* $\mathcal{D}$. Because a dialogue $\tilde{d}$ is an ordered sequence of instantiated locutions, we may consider the sequence of commitment subsets of Φ which arise from this particular sequence as a set of state transitions:

$$\emptyset \cup \{\Phi_{\mathcal{D}}(L_{d,t}(\theta_t)) \,|\, t = 1, 2, \dots\}$$

where each $L_{d,t}(\theta_t)$ is the (instantiated) t-th utterance in dialogue $\tilde{d}$. We append the empty set at the start of this sequence to represent the state of the commitments prior to utterance of the first locution in any dialogue. This means that all dialogues are assumed to commence with the same initial state.

We now have the means by which to identify two dialogues and two dialogue protocols in a semantic sense. In doing so, we are motivated by semantic notions from the theory of programming languages [13]. For example, we may consider two protocols as equivalent if any state transition achievable in one is also achievable in the other, a property known as bisimulation [23]:[7]

Definition 3. ***(Bisimulation Equivalence)*** *For any positive integers j and k, suppose that two partial dialogues $\tilde{d}_j$ and $\tilde{e}_k$ conducted under protocols $\mathcal{D}$ and $\mathcal{E}$ respectively have respective state transitions $\Phi_{\mathcal{D}}(L_{d,j}(\delta_j))$ and $\Phi_{\mathcal{E}}(L_{e,k}(\theta_k))$ such that*

$$\Phi_{\mathcal{D}}(L_{d,j}(\delta_j)) = \Phi_{\mathcal{E}}(L_{e,k}(\theta_k)).$$

Then $\mathcal{D}$ and $\mathcal{E}$ are bisimulation equivalent *if, for any instantiated locution $L_{d,j+1}(\delta_{j+1})$ valid under $\mathcal{D}$, there is an instantiated locution $L_{e,k+1}(\theta_{k+1})$ valid under $\mathcal{D}$ such that*

$$\Phi_{\mathcal{D}}(L_{d,j+1}(\delta_{j+1})) = \Phi_{\mathcal{E}}(L_{e,k+1}(\theta_{k+1}))$$

and conversely.

In other words, bisimulation equivalence says that any transition in commitment states achievable under one protocol by uttering a single instantiated locution can also be achieved under the other using only one instantiated locution. Note that the locutions and the topics with which they are instantiated may differ in the two protocols.

Many protocols permit participants to retract prior utterances. If so, then not all the beliefs expressed or action-commitments incurred during the course of a terminating dialogue may still be current at the end of that dialogue. We therefore distinguish the particular subset of Φ consisting of those beliefs or action-commitments made in the

[7] Strictly, the equivalence defined here is strong bisimulation [23, Chapter 4].

course of a dialogue which are still standing at the normal termination of the dialogue. For a terminating dialogue $\tilde{d}$ conducted under a protocol $\mathcal{D}$, we denote this set by $\Phi_{\mathcal{D},\tilde{d}}$, and we call it the *final commitment-set of* $\tilde{d}$ *under* $\mathcal{D}$. Note that this set may be empty. We therefore have available another notion of protocol equivalence:

Definition 4. ***(Final-State Equivalence)*** *Two protocols* $\mathcal{D}$ *and* $\mathcal{E}$ *are* final-state equivalent *if, for any terminating dialogue* $\tilde{d}$ *conducted under protocol* $\mathcal{D}$*, there is a terminating dialogue* $\tilde{e}$ *conducted under protocol* $\mathcal{E}$ *such that* $\Phi_{\mathcal{D},\tilde{d}} = \Phi_{\mathcal{E},\tilde{e}}$*, and conversely.*

This definition ignores the length of dialogues under each protocol. It would be possible for a dialogue under one protocol to terminate after five utterances (say) and to achieve an outcome for which a dialogue under the second protocol would require 500 locutions. So, we might wish to modify the previous definition as follows:

Definition 5. ***(Equal-length Final-state Equivalence)*** *Two protocols* $\mathcal{D}$ *and* $\mathcal{E}$ *are* equal-length final-state equivalent *if, for any terminating dialogue* $\tilde{d}$ *conducted under protocol* $\mathcal{D}$*, there is a terminating dialogue* $\tilde{e}$ *conducted under protocol* $\mathcal{E}$ *and comprising the same number of utterances as* $\tilde{d}$*, such that* $\Phi_{\mathcal{D},\tilde{d}} = \Phi_{\mathcal{E},\tilde{e}}$*, and conversely.*

For most applications, however, this definition may be too strict. Ideally, we desire a notion of final-state equivalence which would permit terminating dialogues under one protocol to be considered equivalent to terminating dialogues under the other protocol when these had the same outcomes and of similar length. For this notion, we would require a precise definition of the word *"similar."* Moreover, it would be desirable to define this notion so that transitivity is maintained, i.e., so that if protocols $\mathcal{D}$ and $\mathcal{E}$ are similar-length operationally equivalent and if $\mathcal{E}$ and $\mathcal{F}$ are similar-length operational equivalent, then so too are $\mathcal{D}$ and $\mathcal{F}$. We achieve this by partitioning time into a sequence of non-overlapping intervals, as follows:

Definition 6. *Let* $(\, x_i \,|\, i = 1, 2, \ldots)$ *be a finite or countably-infinite sequence of strictly increasing non-negative real numbers, with the first element being* $x_1 = 0$*. In the case where the sequence is finite with* n *elements, assume that* ∞ *is appended to the sequence as the* $n+1$*-th element,* x_{n+1}*. A* ***time partition*** $\mathcal{T}$ *is a collection of closed-open subsets* $\{\, T_i \,|\, i = 1, 2, \ldots\}$ *of the non-negative real numbers* $[0, \infty)$*, such that each* $T_i = [x_i, x_{i+1})$*. If* $\mathcal{T} = [0, \infty)$*, we say it is a* degenerate *time-partition.*

We now use this idea of a partition of time to define a notion of similarity of length for two dialogues. Essentially, two terminating dialogues are said to be of similar length when they both end in the same element of the partition.

Definition 7. ***($\mathcal{T}$-Similar Final-state Equivalence)*** *Let* $\mathcal{T}$ *be any time partition. Two protocols* $\mathcal{D}$ *and* $\mathcal{E}$ *are* $\mathcal{T}$-similar final-state equivalent *with respect to* $\mathcal{T}$ *if, for any terminating dialogue* $\tilde{d}$ *conducted under protocol* $\mathcal{D}$*, there is a terminating dialogue* $\tilde{e}$ *conducted under protocol* $\mathcal{E}$*, such that* $\Phi_{\mathcal{D},\tilde{d}} = \Phi_{\mathcal{E},\tilde{e}}$*, and such that the final utterance of* $\tilde{e}$ *occurs in the same element of time partition* $\mathcal{T}$ *as the final utterance of* $\tilde{d}$*, and conversely.*

It is clear that this notion of equivalence is transitive. Moreover, it can be readily seen that Final-state Equivalence and Equal-length Final-state Equivalence are special cases of $\mathcal{T}$-Similar Final-state Equivalence. In the first case, the partition is the degenerate case of the whole non-negative real line: $\mathcal{T} = [0, \infty)$. In the second case, because dialogue utterances occur only at integer time-points, the relevant partitions are those where each element of the partition includes precisely one integer, for example:

$$\mathcal{T} = [0, 0.5) \cup \bigcup_{k=1}^{\infty} [k - 0.5, k + 0.5).$$

As a final comment regarding these definitions, we note that recent work in abstract concurrency theory has argued that sequential behavior is distinguished from concurrent behavior because the former synchronizes information flows and time, while the latter allows these to evolve independently of one another [25]. If we allow the number of locutions in a dialogue to be a surrogate for time, then we can see that our definition of Final-state Equivalence treats time and information flows as completely independent, since the numbers of locutions in the dialogues under each pair of protocols is not mentioned in Definition 4. In contrast, non-degenerate $\mathcal{T}$-Similar Final-state Equivalence — i.e., all cases where $\mathcal{T} \neq [0, \infty)$ — attempts to re-couple time and information-flows in the pairing of dialogues under the two protocols being considered. Protocols deemed $\mathcal{T}$-similar final-state equivalent do not allow their respective information-flows (in the form of their final commitment sets) and the time taken to achieve these information flows to evolve independently: whatever the link between time and information-flow in any terminating dialogue under one protocol is preserved in the paired dialogue under the other protocol.

4.2 Comparison of Equivalences

We now consider the relationships between these various types of equivalence. We write $\Delta(\Pi)$ to denote the class of all protocols, and $\Delta(P)$ to denote the class of all pairs of protocols $\langle \mathcal{D}, \mathcal{E} \rangle$ where $\mathcal{D}, \mathcal{E} \in \Delta(\Pi)$. Then we write $\Delta(P_{\text{syn}})$ to denote the class of all pairs of protocols $\mathcal{D}, \mathcal{E} \in \Delta(\Pi)$ such that $\mathcal{D}$ and $\mathcal{E}$ are syntactically equivalent, and $\Delta(P_{\text{bi}})$, $\Delta(P_{\text{fin}})$, and $\Delta(P_{\text{eq}})$ to denote the classes of pairs of protocols which are bisimulation equivalent, final-state equivalent, and equal-length final-state equivalent respectively. Moreover, we write $\Delta(P_{\text{sim}})$ to denote the class of all pairs of protocols $\mathcal{D}, \mathcal{E} \in \Delta(\Pi)$ such that there exists a non-degenerate time partition $\mathcal{T}$ for which $\mathcal{D}$ and $\mathcal{E}$ are $\mathcal{T}$-similar final-state equivalent. Call these five classes the *equivalence partitions* of $\Delta(\Pi)$. Then we have the following results:

Proposition 1. *The following set inclusions hold, and each inclusion is strict:*

$$\Delta(P_{\text{syn}}) \subset \Delta(P_{\text{bi}})$$

and

$$\Delta(P_{\text{eq}}) \subset \Delta(P_{\text{sim}}) \subset \Delta(P_{\text{fin}}).$$

Proof. Straightforward from the definitions. □

This proposition says that the class of syntactically-equivalent protocol pairs is a proper subset of the class of bisimulation equivalent protocol pairs. Likewise, the class of equal-length final-state equivalent protocol pairs is a proper subset of the class of non-degenerate $\mathcal{T}$-similar final-state equivalent pairs, which is in turn a proper subset of the class of final-state equivalent protocol pairs.

However, the relationship between bisimulation equivalence and the various forms of final-state equivalence is not straightforward, as the following two results reveal.

Proposition 2. *There exist protocols $\mathcal{D}$ and $\mathcal{E}$ which are bisimulation equivalent but not final-state equivalent.*

Proof. We proceed by demonstrating two such protocols. Consider a protocol $\mathcal{D}$, which contains just one locution, *do*(P_i, θ), a locution which expresses a promise by agent P_i to undertake the action represented by θ, for $\Phi = \{\theta\}$, and for $i = 1, 2$. Further suppose that protocol $\mathcal{D}$ has one rule, a termination rule, which causes the dialogue to terminate normally after any three successive utterances of locution *do(.)*. Thus, there are $2^3 = 8$ terminating dialogues under $\mathcal{D}$, and each has the form:

Utterance 1: *do*(P_i, θ)
Utterance 2: *do*(P_j, θ)
Utterance 3: *do*(P_k, θ)

for possibly-identical agents P_i, P_j and P_k. The final commitment set for each dialogue is either $\{\ (P_1, \theta),\ (P_2, \theta)\ \}$, in the case where both agents have uttered something in the dialogue, or simply $\{\ (P_i, \theta)\ \}$, in the case where all three utterances were made by agent P_i, for $i = 1$ or 2.

Now consider a second protocol $\mathcal{E}$ which has the same locution set as protocol $\mathcal{D}$, but has no termination rules. All dialogues under $\mathcal{E}$ are infinite, and so partial dialogues may have any finite length. However, because Φ consists of a single element, and because there are only two agents, the commitment set for a partial dialogue at any time after commencement has either one or two elements, namely either or both of (P_1, θ) or (P_2, θ). Because there are no retraction locutions, elements in the commitment set cannot be removed once inserted. For protocol $\mathcal{D}$, all partial dialogues have length either one or two locutions. Because the two protocols have the same locutions, it can be readily seen that any transition in commitment states in a partial dialogue achievable under one protocol by uttering a single instantiated locution can also be achieved under the other using the same instantiated locution, and so the two protocols are bisimulation equivalent.

However, the two protocols are not final-state equivalent, since $\mathcal{E}$ has no terminating dialogues. Thus, there do not exist terminating dialogues under $\mathcal{E}$ which can be matched to each of the eight terminating dialogues under $\mathcal{D}$. $\square$

We can also construct a counter-example in the reverse direction, as follows.

Proposition 3. *There exist protocols $\mathcal{D}$ and $\mathcal{E}$ which are equal-length final-state equivalent but not bisimulation equivalent.*

Proof. Again, we demonstrate two such protocols. Consider a protocol $\mathcal{D}$, which contains just one locution, *do*(P_i, θ), a locution which expresses a promise by agent P_i to undertake the action represented by θ, for $\theta \in \Phi$. Here, the set Φ may have more than one element. Further suppose that protocol $\mathcal{D}$ has one rule, a termination rule, which causes the dialogue to terminate normally after any two successive utterances of locution *do(.)*. Thus, all terminating dialogues under $\mathcal{D}$ have the form, for agents P_i, P_j (possibly identical) and $\theta, \delta \in \Phi$:

Utterance 1: *do*(P_i, θ)
Utterance 2: *do*(P_j, δ)

The final commitment set for this dialogue is $\{\theta, \delta\}$.

Now consider a second protocol $\mathcal{E}$ which also contains the locution *do(.)*, with the same syntax. But suppose $\mathcal{E}$ also has a second locution *undo*(P_i, θ), which retracts any prior promise by agent P_i to undertake the action represented by θ. Thus, the dialogue sequence:

Utterance 1: *do*(P_i, θ)
Utterance 2: *undo*(P_i, θ)
.
.

generates the following sequence of commitment states:

$$\emptyset, \{\theta\}, \emptyset, \ldots$$

Next, assume that protocol $\mathcal{E}$ has three combination rules, the first of which states that a valid dialogue must commence with an instantiated utterance of the locution *do(.)*. The second combination rule says that this utterance may be followed either by another instantiated utterance of *do(.)* or by an instantiated utterance of *undo(.)*. The third rule says that subsequent utterances may be instantiations of either locution, subject only to the termination rule. Finally, we assume that $\mathcal{E}$ has one termination rule, which causes a dialogue to terminate normally only in the case of dialogues containing no *undo(.)* locutions, with this termination occurring after two successive utterances of locution *do(.)*.

It can be seen that all terminating dialogues under both protocols $\mathcal{D}$ and $\mathcal{E}$ have the same form, namely:

Utterance 1: *do*(P_i, θ)
Utterance 2: *do*(P_j, δ)

Under both protocols, these dialogues are the same length and lead to the same final commitment set: $\{\theta, \delta\}$. Thus, the two protocols are equal-length final-state equivalent. However, protocol $\mathcal{E}$ contains a commitment state transition which cannot be simulated by any locution in $\mathcal{D}$, namely that effected by the execution of the *undo(.)* locution in the following dialogue sequence:

Utterance k: *do*(P_i, θ)
Utterance $k+1$: *undo*(P_i, θ)

Thus, the two protocols are not bisimulation equivalent. $\square$

It is worth noting that it may be possible to represent similar-length dialogues using order of magnitude reasoning methods, such as those developed in the qualitative physics area of Artificial Intelligence [37]. For example, the system FOG of Olivier Raiman [26], defines three operators to represent the relative values of two physical variables: one variable is negligible relative to the other; their difference is negligible; and both variables are the same size and order of magnitude. Raiman has defined axioms for these three operators, and given the FOG system a semantics based on the calculus of infinitesimals [28].

5 Conclusions

Dialogue game protocols have recently been proposed as the basis for interaction between autonomous agents in a number of situations. As these proposals proliferate, potential protocol users will require guidance in selecting protocols for specified tasks and in choosing between different protocols suitable for the same task. In this paper, we have taken some preliminary steps towards a formal theory of protocols capable of providing such guidance. Building on earlier work classifying dialogues and locutions, we have identified several dimensions by which protocols may be compared, including: the rules which comprise a protocol; the length of dialogues conducted according to a protocol; and the commitments incurred by participants in the course of a dialogue. Our classification of locutions, presented here for the first time, allows for statements of belief about factual matters, and for requests for and promises of actions. Because these locutions connect to external reality (descriptions of the world, and actions in that world) we were able to consider dialogues from the perspective of their semantic effects.

With these dimensions we were able to define several reasonable notions of equivalence of protocols, and to study their relationships to one another. These notions included: syntactic equivalence, where two protocols have identical locutions and rules; bisimulation equivalence, where any semantic transition able to be effected under one protocol can also be effected under the other; and several versions of final-state equivalence, where any final state achievable by a terminating dialogue under one protocol can also be achieved by a terminating dialogue under the other. The various notions of final-state equivalence differ according to whether the matched dialogues are required to have the same, or similar, numbers of locutions, or not. We believe these five notions of equivalence are appropriate for protocols with only terminating dialogues. More work, however, is required on understanding equivalence where one or both protocols permits non-terminating dialogues; this may lead us, in future work, to refine these definitions.

Although the work in this paper is preliminary, we hope it will lead to a complete theory of dialogue-game protocol equivalence for agent communications protocols, and thus provide guidance to protocol designers and users. To our knowledge, these issues have not previously been considered in the agent communications languages community. However, there is much to be done before these initial ideas will comprise a complete theory. Firstly, although we have drawn on notions of equivalence from the theory of programming languages and concurrency theory, there are other notions we could also borrow, such as weak equivalence or congruency [23]. To this end, it may be valuable to further explore the relationships, if any, between interaction protocols and process algebras. Secondly, our definitions of equivalence abstracted away from the details of dialogue game protocols, of dialogues and of locutions presented in Sections 2, 3 and 4. It would be interesting, therefore, to explore notions of equivalence which pertain to specific types of dialogue-games. Thirdly, in developing multiple notions of equivalence it may be useful to articulate desirable properties of such notions, in the same way that two of us have recently identified desirable properties of protocols themselves [21]. In discussing similar-length equivalence in Section 5.1, we mentioned one of these, namely that we believe that protocol-equivalence should be transitive.

These various lines of inquiry may be facilitated by the development of a mathematical language in which to represent protocols, along with a denotational semantics for them [13]. We are currently exploring a category-theoretic semantics for interaction protocols, under which protocols are represented by objects in an appropriate category of protocols. Studying this category and its properties will, we hope, allow us to better understand the space of protocols, and the landscape of this space. For example, we would hope eventually to achieve a denotational characterization of our different notions of equivalence of protocols. Once achieved, we would then seek to identify the best protocol or protocols *within* each equivalence class, according to some reasonable criteria. In addition, for those protocols which are not equivalent, a quantitative measure of their difference would provide guidance to protocol users and designers.

Acknowledgments An earlier version of this work was presented at the Workshop on *Agent Communications Languages and Conversational Policies*, held in Bologna, Italy, in July 2002, as part of the First International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2002). We thank the anonymous referees and the participants in the workshop for their comments. We also thank Rogier van Eijk and Michael Wooldridge for discussions on these topics.

References

1. 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.
2. L. Amgoud and S. Parsons. Agent dialogues with conflicting preferences. In J.-J. Meyer and M. Tambe, editors, *Proceedings of the Eighth International Workshop on Agent Theories, Architectures and Languages (ATAL 2001)*, pages 1–15, 2001.
3. L. Amgoud and S. Parsons. Les dialogues entre agents avec des préférences conflictuelles. In B. Chaib-draa and P. Enjalbert, editors, *Proceedings of the 1 eres Journées Francophones des Modèles Formels de l'Interaction*, pages 3–14, 2001.
4. L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proceedings of the Fourteenth European Conference on Artificial Intelligence (ECAI 2000)*, pages 338–342, Berlin, Germany, 2000. IOS Press.
5. Aristotle. *Topics*. Clarendon Press, Oxford, UK, 1928. (W. D. Ross, Editor).
6. M. Barbuceanu and M. S. Fox. COOL: A language for describing co-ordination in multi agent systems. In V. Lesser, editor, *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, pages 17–24, Cambridge, MA, USA, 1995. MIT Press.
7. T. J. M. Bench-Capon, P. E. Dunne, and P. H. Leng. Interacting with knowledge-based systems through dialogue games. In *Proceedings of the Eleventh International Conference on Expert Systems and Applications*, pages 123–140, Avignon, France, 1991.
8. F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Agent theory for team formation by dialogue. In C. Castelfranchi and Y. Lespérance, editors, *Intelligent Agents VII: Proceedings of the Seventh International Workshop on Agent Theories, Architectures, and Languages (ATAL 2000)*, Lecture Notes in Artificial Intelligence 1986, pages 150–166, Berlin, Germany, 2000. Springer.

9. F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Creating collective intention through dialogue. *Logic Journal of the IGPL*, 9(2):305–319, 2001.
10. A. Finkelstein and H. Fuks. Multi-party specification. In *Proceedings of the Fifth International Workshop on Software Specification and Design*, Pittsburgh, PA, USA, 1989. ACM Sigsoft Engineering Notes.
11. FIPA. Communicative Act Library Specification. Technical Report XC00037H, Foundation for Intelligent Physical Agents, 10 August 2001.
12. M. Greaves, H. Holmback, and J. Bradshaw. What is a conversation policy? In F. Dignum and M. Greaves, editors, *Issues in Agent Communication*, Lecture Notes in Artificial Intelligence 1916, pages 118–131. Springer, Berlin, Germany, 2000.
13. C. A. Gunter. *Semantics of Programming Languages: Structures and Techniques*. MIT Press, Cambridge, MA, USA, 1992.
14. J. Habermas. *The Theory of Communicative Action: Volume 1: Reason and the Rationalization of Society*. Heinemann, London, UK, 1984. Translation by T. McCarthy of: *Theorie des Kommunikativen Handelns, Band I, Handlungsrationalitat und gesellschaftliche Rationalisierung.* Suhrkamp, Frankfurt, Germany. 1981.
15. C. L. Hamblin. *Fallacies*. Methuen, London, UK, 1970.
16. 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 Study of Argumentation (OSSA 2001)*, Windsor, Ontario, Canada, 2001. OSSA.
17. P. McBurney and S. Parsons. Chance discovery using dialectical argumentation. In T. Terano, T. Nishida, A. Namatame, S. Tsumoto, Y. Ohsawa, and T. Washio, editors, *New Frontiers in Artificial Intelligence: Joint JSAI 2001 Workshop Post Proceedings*, Lecture Notes in Artificial Intelligence 2253, pages 414–424. Springer, Berlin, Germany, 2001.
18. P. McBurney and S. Parsons. Representing epistemic uncertainty by means of dialectical argumentation. *Annals of Mathematics and AI*, 32(1–4):125–169, 2001.
19. P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11(3):315–334, 2002.
20. P. McBurney and S. Parsons. Dialogue game protocols. In M.-P. Huget, editor, *Agent Communications Languages*, Lecture Notes in Artificial Intelligence, Berlin, Germany, 2003. Springer. *This volume*.
21. P. McBurney, S. Parsons, and M. Wooldridge. Desiderata for agent argumentation protocols. In C. Castelfranchi and W. L. Johnson, editors, *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2002), Bologna, Italy*, pages 402–409, New York City, NY, USA, 2002. ACM Press.
22. 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*, 2003. *In press*.
23. R. Milner. *Communication and Concurrency*. International Series in Computer Science. Prentice Hall, Hemel Hempstead, UK, 1989.
24. H. Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In M. Ojeda-Aciego, M. I. P. de Guzman, G. Brewka, and L. M. Pereira, editors, *Proceedings of the Seventh European Workshop on Logic in Artificial Intelligence (JELIA-2000)*, Lecture Notes in Artificial Intelligence 1919, pages 224–238, Berlin, Germany, 2000. Springer.
25. V. R. Pratt. Transition and cancellation in concurrency and branching time. *Mathematical Structures in Computer Science*, 2002. *Forthcoming*.
26. O. Raiman. Order of magnitude reasoning. In *Proceedings of the Fifth National Conference on Artificial Intelligence (AAAI-86)*, pages 100–104, San Mateo, CA, USA, 1986. Morgan Kaufmann. Reprinted in [37].

27. C. Reed. Dialogue frames in agent communications. In Y. Demazeau, editor, *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS-98)*, pages 246–253. IEEE Press, 1998.
28. A. Robinson. *Non-Standard Analysis*. North-Holland, Amsterdam, The Netherlands, 1966.
29. J. A. Rodríguez, F. J. Martin, P. Noriega, P. Garcia, and C. Sierra. Towards a test-bed for trading agents in electronic auction markets. *AI Communications*, 11(1):5–19, 1998.
30. 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.
31. J. Searle. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, UK, 1969.
32. C. E. Shannon. The mathematical theory of communication. In C. E. Shannon and W. Weaver, editors, *The Mathematical Theory of Communication*, pages 29–125. University of Illinois Press, Chicago, IL, USA, 1963. Originally published in the *Bell System Technical Journal*, October and November 1948.
33. M. P. Singh. A conceptual analysis of commitments in multiagent systems. Technical Report 96-09, Department of Computer Science, North Carolina State University, 1996.
34. M. P. Singh. An ontology for commitments in multiagent systems: toward a unification of normative concepts. *Artificial Intelligence and Law*, 7:97–113, 1999.
35. M. P. Singh. A social semantics for agent communications languages. In F. Dignum, B. Chaib-draa, and H. Weigand, editors, *Proceedings of the Workshop on Agent Communication Languages, International Joint Conference on Artificial Intelligence (IJCAI-99)*, Berlin, Germany, 2000. Springer.
36. D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany, NY, USA, 1995.
37. D. S. Weld and J. de Kleer, editors. *Readings in Qualitative Reasoning about Physical Systems*. Morgan Kaufmann, San Mateo, CA, USA, 1990.
38. M. J. Wooldridge. *Introduction to Multiagent Systems*. John Wiley and Sons, New York, NY, USA, 2002.