A Mathematical Model of Dialog

Mark W. Johnson [1]

Department of Mathematics
Pennsylvania State University
Altoona PA 16601-3760 USA

Peter McBurney [2]

Department of Computer Science
University of Liverpool
Liverpool L69 3BX UK

Simon Parsons [3]

Department of Computer and Information Science
Brooklyn College
2900 Bedford Avenue
Brooklyn NY 11210 USA

Abstract

Computer Science is currently undergoing a paradigm shift, from viewing computer systems as isolated programs to viewing them as dynamic multi-agent societies. Evidence of this shift is the significant effort devoted recently to the design and implementation of languages and protocols for communications and interaction between software agents. Despite this effort, no formal mathematical theory of agent interaction languages and protocols yet exists. We argue that such a theory needs to account for the semantics of agent interaction, and propose the first mathematical theory which does this. Our framework incorporates category-theoretic entities for the utterances made in an agent dialog and for the commitments incurred by those utterances, together with maps between these.

Key words: agent communications, auctions, category theory, dialogue games, FIPA ACL, interaction protocols, multi-agent systems

---

[1] Email: `mwj3@psu.edu`
[2] Email: `p.j.mcburney@csc.liv.ac.uk`
[3] Email: `parsons@sci.brooklyn.cuny.edu`

1 Introduction

The rise of the Internet, ambient computing, ad-hoc networks and virtual communities have led to a paradigm shift in how we view computer systems and computation [36]. Instead of computer systems being viewed simply as programs which execute some pre-determined method, a better analogy is to view systems as societies of interacting and autonomous entities, or "agents", who combine together as and when necessary to achieve possibly-conflicting individual objectives. This agent-oriented perspective has become influential within computer science over the last decade, and has made connections with prior work in biology (e.g., ecology, evolutionary theory), physics (statistical mechanics), economics (game theory) and sociology (organization theory) [34].

Designing a multi-agent computational system typically means specifying the capabilities and roles of the agents comprising the system, and their means of interaction. Accordingly, considerable research and development effort has been devoted to the design of languages and protocols for autonomous software agents to communicate with one another. The most widely-known language is FIPA's Agent Communications Language [7], which is perhaps the only real standard in this area. [4] FIPA ACL defines 22 locutions, or speech acts, which may be uttered by agents in an interaction in any order, in the same way as humans may freely utter sentences from a human language. Because such freedom leads to a state-space explosion in any realistic application, recent attention has been given to the design of interaction protocols which limit (to a greater or lesser extent) the freedom of agents to make utterances in any order.

The most widespread approach to the design of agent interaction protocols has drawn on dialog games from the philosophy of argumentation, which date at least to Aristotle [4] and which were revived in modern times by Charles Hamblin [9]. They have a structural resemblance to the games of economic game theory [25] and to the two-party games of model theory [15,10]. Agent interaction protocols have been articulated for many different types of dialogue, for example, for dialogs involving Information-Seeking, e.g. [2]; mutual Inquiry [19]; Persuasion [28]; Negotiation over the division of some scarce resource [3]; and Deliberations over what action to take in some circumstance [18]. See [21] for a review of recent work on agent dialogue-game protocols.

In all this work, it is assumed that the agents who enter multi-agent interactions do so for a purpose, although not necessarily a benign or unselfish one. In other words, their behavior is intentional, and so the expected and actual outcome(s) of an interaction are important in understanding it. Thus, any mathematical theory of protocols for such interactions needs to account for the semantics of the interaction, and perhaps also for the semantics of the utterances which comprise the interaction. In seeking such a theory, an obvious starting point would be Claude Shannon's theory of communication [30].

---

[4] This is despite the many problems of FIPA ACL [24,26].

But Shannon, perhaps reflecting his career in a telecommunications company, explicitly ignores the semantics of messages:

> "*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.*" [30, p. 31].

By contrast, for developers and users of agent systems, dealing with the semantics of messages and protocols is an essential part of the "engineering problem" [33]. There are in fact several different functions that a semantics for an agent communications language or dialog protocol may be required to serve:

- To provide a shared understanding to agents participating in a communicative interaction of the meaning of individual utterances, of sequences of utterances, and of dialogs.
- To provide a shared understanding to designers of agent protocols and to the designers (who may be different) of agents using those protocols of the meaning of individual utterances, of sequences of utterances, and of dialogs.
- To provide a means by which the properties of languages and protocols may be studied formally and with rigor, either alone or in comparison with other languages or protocols.
- To provide a means by which languages and protocols may be readily implemented.

Drawing on the programming language semantics literature, van Eijk [6] identified three generic types of semantics for agent communications languages. An **axiomatic** semantics defines each locution of a communications language or protocol in terms of the pre-conditions which must exist before the locution can be uttered, and possibly also the post-conditions which apply following its utterance. For example, the semantic language, *SL*, for the locutions of the FIPA Agent Communications Language, is an axiomatic semantics of the speech acts of the language, defined in terms of the beliefs, desires and intentions of participating agents [7]. Similarly, the semantics defined for many dialog game protocols for agent interaction, e.g., [2], are also axiomatic semantics.

A second type of semantics, an **operational semantics**, considers the dialog locutions as instructions which operate successively on the states of some abstract machine. Here, the semantics defines the locutions in terms of the transitions they effect on the states of this machine. Operational semantics have recently been defined for some agent dialog protocols, e.g., [17]. Third, in **denotational semantics**, each element of the language syntax is assigned a relationship to an abstract mathematical entity, its denotation. Perhaps the first example of a denotational semantics for a dialog protocol was the possible-worlds semantics for question-response interactions defined by Ham-

blin in 1956 [8]. Although possible-worlds and other denotational semantics have a long subsequent history in mathematical linguistics, only recently have denotational semantics been defined for agent dialog protocols. For instance, [22] presents a category-theoretic semantics for a broad class of deliberation dialog protocols, and uses this semantics to prove properties of dialogs conducted under these protocols.

These efforts at defining language and protocol semantics have focused primarily on individual languages or protocols, or on limited classes of protocols. We know of no effort underway to define a semantics for *all* agent interaction protocols. In other words, there is as yet no mathematical theory of agent interaction protocols, in the same way that the lattices of possible worlds semantics provides an algebraic theory for modal logical languages [27]. Such a theory would, we hope, provide a formal means to compare one protocol with another, to identify when two protocols are the same (in some sense or other), and to choose between protocols. Such a theory is the aim of our work.

This paper presents the first mathematical theory of agent interaction protocols, comprising a categorical semantics for a very broad class of protocols. We consider protocols which can be defined explicitly, and abstract away from the type of protocol and the nature of the interaction outcomes. Section 2 of the paper describes our view of Agent Interaction Protocols, Section 3 presents our semantics, while Section 4 presents some mathematical implications of the framework. In Section 5, we present examples of how interaction protocols may be represented in the framework, to illustrate its expressive power. We end the paper with a discussion of future work in Section 6.

It is worth noting that our problem domain and objectives differ from efforts currently underway to develop a semantics for interaction of computational processes in general, such as the work on game semantics [1]. As mentioned above, our domain involves interactions between purposeful agents, each entering a multi-agent dialog with the intention of achieving some goal. Accordingly, the outcomes (both partial and final) of agent interactions are important to any semantic theory, and to the design, engineering and control of the interactions. It is not clear to us that the game-semantics-of-interaction community has considered these issues as prominently as required by the agents community. On the other hand, the outcomes in agent dialogs are considerably more general than the real-valued monetary pay-offs typically assumed in economic game theory [25]. An abstraction of such payoffs is needed for any semantic theory of agent interactions. [5]

---

[5] The only work known to us in economic theory which abstracts from real-valued spaces is [31], but this uses category theory to prove a result about real spaces.

2 Agent Interaction Protocols

The syntactical form of an agent interaction protocol may be defined explicitly by specifying a number of elements [20], as follows:

Commencement Rules: Rules which define the circumstances under which the dialog commences. Typically, the Commencement Rules of a protocol refer to states prior to or outside the dialogue, and so are not modelled within it. We will not consider these further in this paper.

A collection of Locutions: Rules which 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 that have been subsequently questioned or contested to justify their assertions. Justifications may involve the presentation of a proof of the proposition or an argument for it. [6]

Combination Rules for the Locutions: 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 p and subsequently the proposition $\neg p$ in the same dialogue, without in the interim having retracted the former assertion. Another example involves argumentative dialogue, where one agent may request another to justify a statement the latter has made; most such protocols require the claimant to respond to such a request immediately after the justification request is made.

A collection of Commitments: Some utterances in a dialog may commit the speaker to some claim or action. A bid in an auction, for example, commits the bidder to purchase the good in question at the price mentioned in the bid, if the bid is accepted. Typically, the assertion of a claim p in a debate is defined as indicating to the other participants some level of commitment to, or support for, the claim. Since [9], formal dialog systems typically establish and maintain public sets of commitments, called *commitment stores*, for each participant; these stores may be non-monotonic, in the sense that participants may also be permitted to retract committed claims, although possibly only under defined circumstances.

Combination Rules for the Commitments: Rules which define the ways in which Commitments may be combined or not. For example, it is usually not permitted for an agent, in the one dialogue, to commit to undertake some action and to subsequently commit not to undertake the same action, without first having withdrawn or cancelled the first commitment. Note that an agent who makes a commitment may not be able to withdraw or modify it without permission from other agents, depending on the rules of the dialog, as in [23].

---

[6] Classifications of locutions have been given, for example, by [5,29].

Locution-Commitment Assignment Rules: An assignment of a commitment or commitments to each locution, in a manner compatible with the relevant combination rules.

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

Thus, a dialog under a protocol defined by a structure such as this consists of an ordered sequence of locutions which is not forbidden by the combination rules for locutions. The commitment associated to a dialog then refers to the combination of the ordered sequence of commitments associated to these locutions. For example, an agent may make an offer at one point in a dialog and later retract this offer, if the protocol permits this. Even if a retraction utterance is permitted, the commitment associated to the initial offer may or may not then be cancelled by the commitment associated to the retraction, depending on the commitment combination rules of the protocol. Retraction of a prior offer may incur a penalty, for example, so that the commitments created by the prior offer still stand.

With respect to commitments, it is worth noting here that more than one notion of *commitment* is present in the literature on dialog games. For example, philosophers of argumentation often treat commitments in a purely dialogical sense, so that they may have no reference to anything beyond the dialogue, e.g., [9]. In contrast, others treat commitments as obligations to (execute, incur or maintain) a course of action [32]. 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 propositional commitment can be seen as a special case of action commitment. Because our primary motivation is the design of interaction protocols between autonomous software agents, 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 view commitments as referring to some objects (physical or virtual) in the world beyond the dialogue. [7]

Of course, this structure does not capture all agent interaction protocols, for example, those which cannot be defined formally or finitely. However, it is sufficiently general to represent protocols of each type commonly seen in human or agent dialogues, such as those defined in the typology of [32]. In the next section, we present a categorial semantics for all protocols definable with this structure.

---

[7] In addition, within multi-agent systems research, *commitment* is sometimes used to refer to persistent intentions. This is not the usage here.

3   The Categorical Framework

We begin our presentation with some explanatory words on category theory [16]. A (small) category is a minimalist mathematical construct which consists of two sets and a system of combination rules. The first set, called the objects of the category, is largely a placeholder. The second set, called the morphisms of the category, consists of a collection of arrows from one object to another. Thus, one might think of a morphism as an arrow with a tail (or source) and a head (or target). Given a pair of arrows, one can try to combine them to form a longer arrow if the head of one lies at the same object as the tail of the other.

In other words, we can separate the collection of all morphisms in our category $\mathcal{C}$ into sets as those with the same head and tail. Then $\mathcal{C}(A, B)$ will denote the set of all arrows $A \to B$ and the composition law is an assignment

$$\mathcal{C}(B, C) \times \mathcal{C}(A, B) \to \mathcal{C}(A, C)$$

of an arrow $gf : A \to C$ to every pair consisting of $f : A \to B$ and $g : B \to C$. The reason for writing gf rather than fg comes from the theory of mathematical functions, but the reader should keep in mind that "time flows from right to left" in this notation. In other words, gf represents first following the arrow f and then following the arrow g.

The first element of our model is to consider the locutions of a protocol as the arrows (or, more properly, morphisms) in a category where the composition law is determined by the combination rules for locutions in that protocol. We will tend to use the symbol $\mathcal{D}$ for this category. [8]

There is one formal complication which arises from our desire to forbid certain combinations of locutions. In order to deal with this problem, we add a new element $*$ to every set $\mathcal{C}(A, B)$ which we think of as an illegal arrow. We also want to say any composition with the illegal arrow on either side is another illegal arrow. The technical terminology for this process is to consider only categories "enriched over pointed sets". Notice, we can now say f followed by g is illegal within the categorical context by the equation $gf = *$. This complication should be viewed as purely formal and will be suppressed whenever this will cause no additional confusion. [9]

One should keep in mind that the composition rule in a category must be associative. This is simply the statement that $(hg)f = h(gf)$, so that all compositions can be formed in whatever order is convenient. In particular, this means that making hgf an illegal combination then implies both that h may not legally follow gf and that hg may not legally follow f. One other key feature of a category is that there is an identity morphism 1_B associated

---

[8]   For those who take the view that a protocol consists solely of locutions and their combination rules, the category $\mathcal{D}$ alone then provides a model for a protocol.
[9]   For more on this construction, see [11].

to each object B of the category. This has the property that $1_B f = f$ for any $f : A \rightarrow B$ and $g 1_B = g$ for any $g : B \rightarrow C$. Now a dialog represents a sequence of composable arrows $(f_n, f_{n-1}, \ldots, f_1)$, where composable simply means the target of f_i and the source of f_{i+1} are the same object. The dialog is illegal precisely when the composition is illegal, $f_n f_{n-1} \ldots f_1 = *$.

Because categories include identity arrows, from each object to itself, we can actually do away with the objects of the category, and simply have a collection of arrows. Thus, the arrows in our model represent the locutions in a dialog protocol, while the objects are simply placeholders, indicating the start-points and end-points of locutions. In addition, in order to determine which dialogs terminate, one may add an "end of dialog" locution, so that terminating dialogs are precisely those which include this locution.

We have not yet dealt with the commitments in any way. As with the locutions, one might build a category of commitments with the commitments as arrows and their combination rules determining the composition law of the category. We will use $\mathcal{O}$ to indicate this category. However, in many cases it seems one would like to consider all commitments as "composable". Mathematically, this corresponds to assuming that there is only one object in $\mathcal{O}$. In this case, $\mathcal{O}$ may be more efficiently described as a monoid, which simply means a set together with a multiplication that may (or may not) have inverses but does contain a unit. The set in question is the set of all morphisms of $\mathcal{O}$, the multiplication is given by the composition law and the unit comes from the identity map of the unique object. For example, the whole numbers under addition form a monoid, with 0 as the unit. As another example, consider the integers under multiplication with 1 as the unit.

The question of the existence of inverses corresponds to the question of which commitments may be retracted without restriction. If all commitments have a retraction, the monoid becomes a group, a mathematical construct which may be more familiar than either monoids or categories. The group associated to the whole numbers under addition is the integers under addition. The group associated to the integers under multiplication is the rational numbers (fractions of integers) under multiplication. In fact, one can always find a smallest group that contains a particular monoid, which would allow us to focus on groups rather than monoids if we prefer. Notice the group would be Abelian (multiplication order is irrelevant) precisely when commitments are all time-independent of one another.

We still have not dealt with the Locution-Commitment Assignment Rules in the protocol structure. This involves some assignment of an arrow in $\mathcal{O}$ to each arrow in $\mathcal{D}$ in a manner compatible with the composition laws in the two categories. The term for an assignment between categories is a functor $F : \mathcal{D} \rightarrow \mathcal{O}$, which associates an object $F(D)$ of $\mathcal{O}$ to each object D of $\mathcal{D}$. Further, associated to each arrow $g : B \rightarrow C$ in $\mathcal{D}$, one has an arrow $F(g) : F(B) \rightarrow F(C)$ in $\mathcal{O}$. Finally, for composable morphisms g and f, one has $F(gf) = F(g)F(f)$ so one can compose and then map to $\mathcal{O}$ or map each

arrow to $\mathcal{O}$ and then compose, with the same results. Once again, we have technical restrictions due to the illegal morphisms, so we would like to force $F(1_B) = 1_{F(B)}$ and $F(*) = *$ as well. This says F must be an enriched functor between the two categories $\mathcal{D}$ and $\mathcal{O}$ which are enriched over pointed sets. Note that enriched category theory is a mature mathematical theory [13], and so we have access to a variety of well-known constructions.

Thus, our model for a protocol consists of a triple:

$$\mathcal{D}, \mathcal{O} \text{ and } F : \mathcal{D} \to \mathcal{O}$$

where $\mathcal{D}$ and $\mathcal{O}$ are categories (enriched over pointed sets) with F an enriched functor.

4 Implications of the Framework

In the theory of categories, there is an obvious category which contains all of the functors $F : \mathcal{D} \to \mathcal{O}$ once we fix $\mathcal{O}$. This would be called the category of pointed categories over $\mathcal{O}$ and could be denoted $\mathrm{Cat}_* / \mathcal{O}$. The morphisms in this category from $F : \mathcal{D} \to \mathcal{O}$ to $G : \mathcal{C} \to \mathcal{O}$ are the (enriched) functors $H : \mathcal{D} \to \mathcal{C}$ which make the following triangle commute:

$$\begin{array}{ccc} \mathcal{D} & \xrightarrow{\ H\ } & \mathcal{C} \\ & {\scriptstyle F}\searrow \quad \swarrow{\scriptstyle G} & \\ & \mathcal{O}. & \end{array} \qquad (1)$$

Recall that a diagram is said to commute if each path through the diagram yields the same result at any point when results can be compared, so this triangle commuting says $GH = F$. This implies there is a natural notion of morphism between the triples associated to protocols with the same commitment category, and they form the category $\mathrm{Cat}_* / \mathcal{O}$.

Among the most basic objects in category theory are categories denoted $[n]$, which contain only a string of n composable morphisms aside from the required identity and illegal maps. For example:

$$[1] \qquad 0 \longrightarrow 1 \qquad\qquad (2)$$

$$[2] \qquad 0 \longrightarrow 1 \longrightarrow 2$$

$$[3] \qquad 0 \longrightarrow 1 \longrightarrow 2 \longrightarrow 3$$

Then $[0]$ simply consists of a single object and its identity and illegal morphisms. The main use for a category $[n]$ is that a functor $G : [n] \to \mathcal{D}$ is

simply a string of composable morphisms. However, even if $\mathcal{D}$ is a category of locutions, the morphism $G(k) \to G(k+1)$, which is an arrow between two objects in $\mathcal{D}$, may correspond to either the identity on the object $G(k)$ or a long string of combinable locutions in $\mathcal{D}$.

Suppose $\mathcal{O}$ has only one object (as in the auction examples considered below) and $H : [n] \to \mathcal{O}$ is a functor. Then H corresponds to the choice of an ordered sequence of commitments and this makes H an object in $\mathrm{Cat}_* / \mathcal{O}$. Thus, we can consider a morphism in $\mathrm{Cat}_* / \mathcal{O}$, which consists of a commutative triangle

$$[n] \xrightarrow{\;\;G\;\;} \mathcal{D} \qquad\qquad (3)$$
$$\underset{H}{\searrow} \quad \underset{F}{\swarrow}$$
$$\mathcal{O}.$$

This corresponds to a dialog in our protocol whose associated commitment sequence is the ordered sequence of commitments associated to H. However, it may be that we have compressed the dialog by looking at a longer dialog and reducing it down to a shortened apparent length by composing certain portions and ignoring intermediate commitments over the excised period. In any event, the functor G chooses a sequence of composable morphisms in $\mathcal{D}$, which corresponds to a dialog in our protocol (although we may be fast-forwarding certain portions of the dialog in some sense). The assumption that the triangle commutes implies the commitments associated to the relevant portions of the dialog must be those associated to H. If we only want to pay attention to the final commitments rather than to an ordered sequence of intermediate commitments, we should simply consider the case $n = 1$. The reader should be aware that we do NOT assume all dialogs begin at the same point in this work.

A natural object to associate to a model of a protocol is the set of all such commutative triangles (graded by n). This is a very natural construction in category theory and corresponds to the simplicial set associated to a category, often called the **nerve** of the category. In terms of protocols, this corresponds to looking at dialogs sorted by their ordered sequence of outcomes, possibly by ignoring intermediate commitments. If we restrict to what are usually called the one-simplices, or setting only $n = 1$, this corresponds to dialogs where we consider only the final commitments. There are a large number of notions of equivalence of simplicial sets and we are currently applying these to the study of protocol equivalence.

5   Some Examples

In this Section, we present some illustrative examples of agent interaction protocols represented in our categorical framework. The basic idea is that the locutions create a directed graph by tiling and then the combination rules impose relations via the composition law.

5.1 FIPA ACL

As mentioned above, FIPA ACL, the Agent Communications Language of the Foundation for Intelligent Physical Agents (FIPA), defines 22 locutions which may be uttered by agents in a dialog in any order [7]. These include locutions to *inform* another agent of the truth of some proposition, or to *request* that some action be undertaken. FIPA ACL does not define any locution combination rules, so that an agent may utter any of the 22 locutions at any point in a dialogue. This means the category $\mathcal{D}$ should be (basepoints added to) the free category on the 22 possible locutions, essentially just a repeated tiling where the tile consists of 22 morphisms with the same source and all targets are different. (See below for a tiling example with only three morphisms.) Free objects are one of the most carefully studied concepts in category theory, sometimes thought of as the most universal constructions. This connection with a free category explains the feeling that many other protocols could be modeled by imposing relations on the FIPA ACL.

For the commitment category in this example, notice that no commitments are defined or associated to locutions in the FIPA ACL, hence there are no combination rules for commitments. In our framework, this may be represented by saying the commitment category should simply be [0] described above. That is, we should instead think of a single outcome which iterates to itself as the outcome associated to each (legal) locution.

To understand the functor $F : \mathcal{D} \rightarrow [0]$, it will suffice to notice that no combination of locutions is forbidden, so the basepoints are an afterthought in this case. In order to represent this, we should think of F as adding basepoints to an ordinary functor from a free category to the category with one object and only the identity morphism. There is only one such functor into such a trivial category, namely the functor which sends all objects to the unique object and all morphisms to the identity morphism. Thus, our functor F will send only the basepoint morphisms to the basepoint morphism of [0] and every other morphism will be sent to the identity morphism in [0].
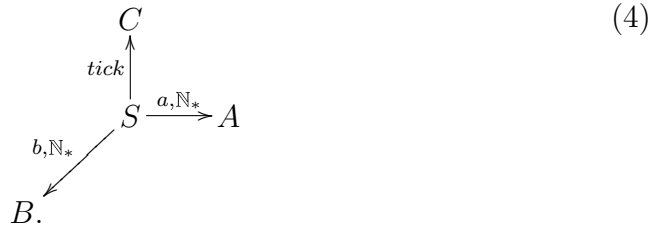
5.2 An English Auction

Perhaps the most widely-used formal interaction protocols are auctions. These are processes by which one or more buyers negotiates the price of some good with one or more sellers [14]. In the most common form of auction, the so-called *English* auction, multiple potential buyers of a single good bid increasingly higher prices to purchase the good from a single seller. The winning bidder is that potential buyer who makes the highest bid, and the amount paid by the buyer is the amount indicated in that highest bid. Each bid may be viewed as an utterance creating a commitment to purchase the item if agreed by the seller. We can represent this process by viewing our category as a tiling, where a single tile is defined by the number of atomic locutions and the set of parameters allowable for each.

For example, suppose a basic auction protocol (for two bidders) consists of three possible utterances:

"Agent a increments the current bid by amount n";

"Agent b increments the current bid by amount m"; and

"The clock ticks with no bid".

When a then increments the current bid, this is an atomic locution, while n is a parameter which would generally be a natural number. However, zero is always the parameter for the clock. (The clock is only included so that the end of the auction is detectable by three consecutive clock ticks.) The basic "tile" would then consist of four objects (one more than the number of atomic locutions), which we will label S, A, B, and C. Then $\mathcal{D}(S, A)$ would be the (pointed) natural numbers (the possible parameters) corresponding to the first locution where a increments the current bid. Similarly, $\mathcal{D}(S, B)$ would be the (pointed) natural numbers corresponding to the locution where b increments the current bid. Finally, $\mathcal{D}(S, C)$ would be two points, one corresponding to the clock tick and the other to the illegal locution. There would be no other morphisms aside from the required identities and illegal morphisms.

A diagram of this tile would be as follows:

$$\begin{array}{c} C \\ \big\uparrow {\scriptstyle tick} \\ {\scriptstyle b,\mathbb{N}_*} \swarrow \ S \xrightarrow{\ a,\mathbb{N}_*\ } A \\ B. \end{array} \qquad (4)$$

Now the point of the tiling idea is that we would think of each of 1, 2 or 3 as a new location for 0. One iteration of this process might yield the following diagram (where new objects are de-emphasized):

$$(5)$$
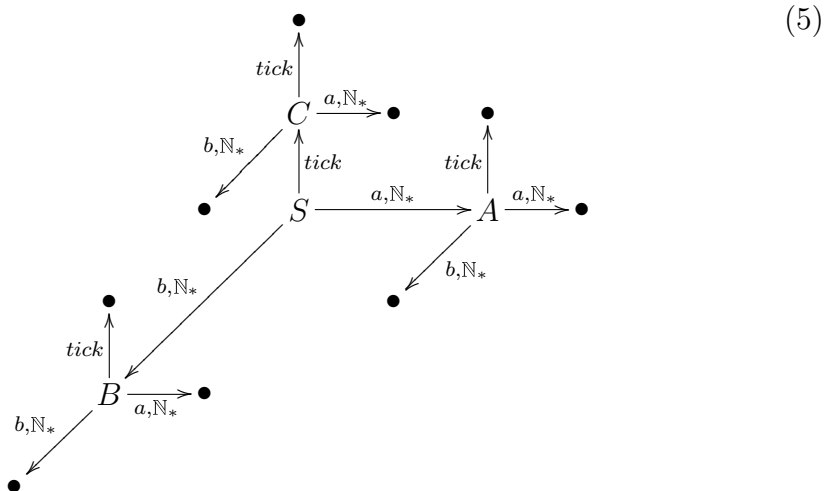
Iterating this procedure yields something like a lattice in $\mathbb{R}^n$, which can be described as a free category (enriched over pointed sets). However, we now

need to introduce the relations inherent in our combination rules for locutions. In the case of our auction example, "a increments the bid by n" followed by "b increments the bid by m" should be viewed as equivalent to "b increments the bid by m+n", for example. Notice we also avoid much of the state-space explosion problem of the FIPA ACL in this case, since any three consecutive ticks ends the dialog, allowing us to impose a height restriction in this diagram.

Our outcome category for the English auction would consist of the pointed whole numbers times a small pointed monoid consisting of 1, a, b and the basepoint, which keeps track of the last real bidder (any tick of the clock would be given the identity in the bidder slot). The functor would simply take "a increments the bid by n" to the pair (a, n) in this notation, so our relation above becomes $(b, m)(a, n) = (b, m + n)$.

6 Conclusions

In this paper, we have presented a mathematical theory of agent interaction protocols which takes explicit account of the semantics of protocols. We do this by representing formally the utterances and commitments in agent dialogs, and the relationships between them. Our model is a categorical one, and it abstracts away from the type of interaction and the nature of the commitments being discussed. It therefore applies to a very broad class of agent interaction protocols, and is also not limited to real-valued monetary transactions. In contrast, prior work on the semantics of agent dialogs has focused on the semantics of individual utterances, as in the semantics of the FIPA ACL [7], or on the semantics of dialogs under only one protocol, as in [17], or a limited class of protocols, as in [22]. Similarly, prior work on parametrizing the space of auction mechanisms, such as [35], does not extend to dialogue game protocols.

This feature of our work helps answer an important question: *Why use category theory?*. Only category theory is sufficiently abstract that we could hope to represent all types of agent interaction protocols. That we were able to present a model of the FIPA ACL, an interaction language defined without explicit commitments, and a model of an auction protocol, in which utterances are usually assumed to incur commitments, shows the potential of this formalism. In addition, a categorical semantics is likely to prove necessary to answering the question: *When are two protocols the same?* In earlier work [12], we identified several distinct notions of *protocol equivalence*, and we are currently representing these different notions in our framework. A mathematical theory of protocols should be able to characterize different types of protocols and identify those which are similar or equivalent. A categorical semantics may also allow us to build new protocols with specific properties. Our future work is devoted to exploring the implications of this framework and applying it to protocol comparisons.

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 International Conference on Multi-Agent Systems (ICMAS 2000)*, pages 31–38, Boston, MA, USA, 2000. IEEE Press.

[3] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proceedings of the European Conference on Artificial Intelligence (ECAI 2000)*, pages 338–342, Berlin, Germany, 2000. IOS Press.

[4] Aristotle. *Topics.* Clarendon Press, Oxford, UK, 1928. (W. D. Ross, Editor).

[5] J. L. Austin. *How To Do Things with Words.* Oxford University Press, Oxford, UK, 1962.

[6] R. M. van Eijk. *Programming Languages for Agent Communications.* PhD thesis, Department of Computer Science, Utrecht University, Utrecht, The Netherlands, 2000.

[7] FIPA. Communicative Act Library Specification. Standard SC00037J, Foundation for Intelligent Physical Agents, 3 December 2002.

[8] C. L. Hamblin. *Language and the Theory of Information.* Ph.D. thesis, Logic and Scientific Method Programme, University of London, London, UK, 1957. Submitted October 1956.

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

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

[11] M. W. Johnson. On pointed enrichments and illegal compositions. Technical Report ULCS-03-010, Department of Computer Science, University of Liverpool, Liverpool, UK, 2003.

[12] M. W. Johnson, P. McBurney, and S. Parsons. When are two protocols the same? In M-P. Huget, editor, *Communication in Multi-Agent Systems: Agent Communication Languages and Conversation Policies*, Lecture Notes in Artificial Intelligence 2650, pages 253–268. Springer, Berlin, 2003.

[13] G. M. Kelly. *Basic Concepts of Enriched Category Theory.* London Mathematical Society Lecture Notes 64. Cambridge University Press, Cambridge, UK, 1982.

[14] V. Krishna. *Auction Theory*. Academic Press, San Diego, CA, USA, 2002.

[15] P. Lorenzen. Ein dialogisches konstruktivitätskriterium. In *Infinitistic Methods: Proc. Symp. Foundations of Mathematics, Warsaw, 2-9 September 1959*, pages 193–200, Warszawa, Poland, 1961. PWN.

[16] S. Mac Lane. *Categories for the Working Mathematician*. Graduate Texts in Mathematics 5. Springer, New York, second edition, 1998.

[17] P. McBurney, R. M. van Eijk, S. Parsons, and L. Amgoud. A dialogue-game protocol for agent purchase negotiations. *Journal of Autonomous Agents and Multi-Agent Systems*, 7(3):235–273, 2003.

[18] P. McBurney, D. Hitchcock, and S. Parsons. The eightfold way of deliberation dialogue. *Intelligent Systems*, 2005. *In press*.

[19] 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.

[20] 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.

[21] P. McBurney and S. Parsons. Dialogue game protocols. In M-P. Huget, editor, *Communication in Multi-Agent Systems: Agent Communication Languages and Conversation Policies*, Lecture Notes in Artificial Intelligence 2650, pages 269–283. Springer, Berlin, 2003.

[22] P. McBurney and S. Parsons. A denotational semantics for deliberation dialogues. In N. R. Jennings, C. Sierra, E. Sonenberg, and M. Tambe, editors, *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS 2004)*, pages 86–93, New York City, 2004. ACM Press.

[23] P. McBurney and S. Parsons. The Posit Spaces Protocol for multi-agent negotiation. In F. Dignum, editor, *Advances in Agent Communication*, Lecture Notes in Artificial Intelligence 2922, pages 364–382. Springer, Berlin, 2004.

[24] 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)*, pages 402–409, New York City, 2002. ACM Press.

[25] M. J. Osborne and A. Rubinstein. *A Course in Game Theory*. MIT Press, Cambridge, MA, USA, 1994.

[26] J. Pitt and A. Mamdani. Some remarks on the semantics of FIPA's Agent Communications Language. *Journal of Autonomous Agents and Multi-Agent Systems*, 2:333–356, 1999.

[27] S. Popkorn. *First Steps in Modal Logic.* Cambridge University Press, Cambridge, UK, 1994.

[28] H. Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In M. Ojeda-Aciego et al., editors, *Proceedings of the Seventh European Confernce on Applications of Logic in Artificial Intelligence (JELIA 2000)*, Lecture Notes in Artificial Intelligence 1919, pages 224–238, Berlin, Germany, 2000. Springer.

[29] J. Searle. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge University Press, Cambridge, UK, 1969.

[30] 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.

[31] H. Sonnenschein. An axiomatic characterization of the price mechanism. *Econometrica*, 42(3):425–434, 1974.

[32] 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.

[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. *Introduction to Multiagent Systems.* Wiley, New York, 2002.

[35] P. R. Wurman, M. P. Wellman, and W. E. Walsh. A parametrization of the auction design space. *Games and Economic Behavior*, 35(1–2):304–338, 2001.

[36] F. Zambonelli and H. v. D. Parunak. Signs of a revolution in computer science and software engineering. In P. Petta et al., editors, *Engineering Societies in the Agents World (ESAW 2002)*, Lecture Notes in Artificial Intelligence 2577, pages 13–28, Berlin, 2003. Springer.