Coordination through Inductive Meaning Negotiation*

Alessandro Agostini
ITC-IRST Trento

April 23, 2002

Abstract

This paper is on negotiation, precisely on the negotiation of *meaning*. We advance and discuss a formal paradigm of coordination and variants thereof, wherein meaning negotiation plays a major role in the process of convergence to a common agreement. Our model engage a kind of pairwise, model-theoretic coordination between knowledge-based agents, eventually able to communicate the *complete & local* meaning of their beliefs by expressions taken from the literals of a common first-order language. We address the question of how the model provides the basis for a computational approach to the motivating problem of coordination by meaning negotiation. We exhibit a computable agent who coordinates with every agent taken from a uniformly computable class.

1 Introduction

Following [JFL+01], automated negotiation research can be considered to deal with three broad topics, namely: (a) *Negotiation Protocols*: the set of "rules" that govern the interaction among the agents in play. (b) *Negotiation Objects*: the range of issues, or attributes, over which agreement must be reached. (c) *Agents' Decision Making Models*: the decision making functionality the agents concerned employ to act in line with the negotiation protocol (a) in order to achieve their objectives (b). References to the literature on negotiation protocols and objects are, for instance, [RZ94, JFL+01, WP00, Syc89] and the references cited there. Either negotiation protocols togheter with negotiation objects or agents' decision making models is the dominant concern. In this paper we are interested in a unique negotiation object: *meaning*.

The negotiation of meaning is an inductive process that is shaped by multiple elements and that affects these elements. Meaning negotiation constantly changes the situations to which it gives meaning and affects all participants. In this process, negotiating meaning entails both language interpretation and action. As our approach suggests, language interpretation is "local," in the sense that it is hidden in individual agents. On the other hand, actions are "global," since they directly

*Electronic correspondence to: agostini@irst.itc.it.

influence the agents' environment. This is true expecially if we think of them as linguistic "acts" of a common communication language. The contrast between the local interpretations of language of agents and the global actions is at the core of a number of interesting problems of coordination and forms the main perspective of the work presented in this paper. In short, even if the agents have a common interest to coordinate, coordination success by meaning negotiating is far from trivial. In fact, the negotiation agents may not be sure that they have made mutually acceptable proposals, since the association between linguistic forms and their meaning, that is the agents' language interpretation, is a local procedure.

In this paper, we propose a new approach to the study of coordination between knowledge-based agents, eventually able to communicate the *local* meaning of their beliefs by expressions taken from a common language. In our approach, meaning negotiation plays an important role, as it seems to be a nice way to overcome the difficult problem of locality of language interpretation. We show why the coordination competence of a special kind of agents cannot be strictly improved (Corollary (15)), and the existence of a computable agent who coordinates with every member of a "uniformly computable" class of computable agents (Theorem (17)). We proceed along the model-theoretic, first-order tradition of *formal learning theory*—say [Kel96, OdJMW97, MO98] and the references cited there, that descends from the pioneering studies on inductive inference developed by Solomonoff, Putman, Gold, Blum, among others (see for instance [JORS99] for a survey and references).

This paper is organized as follows. In the next section we present an intuitive game-theoretic picture of our basic paradigm of coordination, followed (Section 3) by the formal development of the model. In Section 4 we illustrate the behavior of a special kind of "self-centered agents," and give a result on their coordination competence. In Section 5 we address the question on how our paradigm provides the basis for a computational approach to the motivating problem of coordination by meaning negotiation. Section 6 reports on some related work. Section 7 is the conclusion.

2 A Game for Meaning Negotiation

The following game is illustrative of our basic model. The game is played between two players. Call them "the seeker" (also: system, learner, server...) and "the source" (also: user, teacher, client...). The following pieces are used.[1]

1. A (countable, decidable) first-order language $\mathcal{L}$, with identity ($\doteq$) and variables v_0, v_1, v_2, ...;
2. for each player $i = 1, 2$, a collection $\mathbf{K}_i$ of countable structures that interpret $\mathcal{L}$, and
3. for each player $i = 1, 2$, a partition P_i of the collection $\mathbf{K}_i$ into cells.

[1]Most part of this section is dealt with systematically in the next section. In what follows, we assume standard concepts of first-order logic.

To illustrate, let the signature of $\mathcal{L}$ be limited to a single binary predicate, say, R. Let the countable structures $\mathbf{K}_1 = \mathbf{K}_2$ be all those that interpret R as a linear order with either a least point or a greatest point, but not both. And let the partitions $P_1 = P_2$ be based on the existence of a least point with respect to R's interpretation. There are only two cells. In one cell are all the linear orders with a least point. In the other cell lie the linear orders with a greatest point.

The source prepares herself for the game as follows. First she chooses one structure $\mathcal{S}$ from one of the two cells of her partition. We call such a structure the source's "world-in-play." Pursuing our illustration, she may choose the structure consisting of the natural numbers N with R interpreted as the usual linear order $\leq$ on N. She then maps the variables of $\mathcal{L}$ onto the domain of $\mathcal{S}$, thereby using the variables as temporary names for all the objects in the countable universe of $\mathcal{S}$. The source is allowed to choose any surjective mapping she pleases. For example, she might map v_i to the number i. An important point to note is that the source's choice of structure and variable-mapping is generally *unknown* to the seeker—this is indeed the distinction of "local" and "global" we mentioned in the introduction. We call the pair $\langle$ structure, variable-mapping $\rangle$ the source's "knowledge-in-play."

Now the game begins. The source starts by selecting a literal that is true in her knowledge-in-play. Following our example, she might choose any of $v_0 R v_1$, $v_7 R v_{22}$, $\neg v_9 R v_8$, $v_1 \doteq v_1$, $v_3 R v_3$, etc. In response to the source's literal, the seeker first chooses a structure from his partition together with a variable-mapping, then selects a literal that is true in his so formed knowledge-in-play. Again, the seeker's choice is hidden to the source. The seeker is allowed to choose any structure and variable-mapping he pleases. For example, he might choose the identical structure $\mathcal{S}$ the source has chosen and map both v_{2i} and v_{2i+1} to the number i. It is then the source's turn again. She chooses a new literal that is true in her knowledge-in-play. So two literals now stand revealed by the source. The seeker responds to the two literals by possibly changing the initial choice of his knowledge-in-play—this would happen, for example, if the seeker realizes that either the structure or the variable-mapping he has chosen is not "sufficiently similar" to the source's structure and variable-mapping, and selecting a literal that is true in his (eventually new) knowledge-in-play. The two players go on like this forever. In each round, the source reveals a literal, and the seeker chooses a knowledge-in-play (possibly identical to the previous step's choice) and plays a literal.

Here are the rules. The source can't change her knowledge-in-play. Importantly, the source can withhold no fact forever. That is, for every literal that is true in the source's knowledge-in-play, there must be a round of the game in which the source reveals it. The seeker wins the game just in case he chooses in the limit (that is, "converges to") a knowledge-in-play that is "sufficiently similar" to the source's knowledge-in-play. This means that his choice stabilizes on a structure in the same cell of the partition from which the source's structure was drawn.

3 The Paradigm

In this section we formalize the foregoing ideas. For doing this, we need some preliminary notation.

Note on Notation. We denote the set $\{0, 1, 2, ...\}$ of natural numbers by N. We denote the usual linearly ordered structure with domain N by ω. Let η be an infinite sequence.[2] For $i \in N$, we write $\eta|_i$ for the proper initial sequence of length i in η. For every finite or infinite sequence ζ of length $n > k$, $k \in N$, we let $\zeta[k]$ denote the finite sequence $\langle \zeta_0 \cdots \zeta_k \rangle$ and ${}_k|\zeta$ denote the sequence obtained from ζ by deleting its first $k + 1$ elements. We write $|\sigma|$ for the length of a finite sequence, and $\emptyset$ for the finite sequence of length zero, Thus, $\zeta[k] = \zeta|_{k+1}$, and ${}_0|\zeta = \emptyset$ if $|\zeta| = 1$. We write σ_i or also $(\sigma)_i$ for the ith element of σ, $0 \leq i < |\sigma|$ Concatenation of sequences is indicated by juxtaposition and we won't distinguish notationally between an element and the corresponding unit sequence. Thus $\alpha\tau$ first element α and tail τ. The set of elements in a (finite, infinite) sequence τ is denoted by $range(\tau)$.

3.1 A First-Order Framework

Let us begin by establishing a first-order framework. We write $\mathcal{L}_{form}$ to denote a first-order language with equality built up from a (countable, decidable) vocabulary (equality symbol $\doteq$) and a countably infinite set $Var = \{v_i \mid i \in N\}$ of variables. We write $\mathcal{L}_{sen}$ to denote the set of *sentences*, that is, the subset of $\mathcal{L}_{form}$ containing no free variables. We write $\mathcal{L}_{basic}$ to denote the set of literals of $\mathcal{L}_{form}$. The members of $\mathcal{L}_{basic}$ are called *basic formulas.*

Our semantic notions are standard. Let $\mathcal{S}$ be a structure that interprets the vocabulary of $\mathcal{L}_{form}$. Let $\models$ denote the model theoretic concept of truth in a structure. Then $\mathcal{S}$ is a *model* of $\Gamma \subseteq \mathcal{L}_{form}$, and Γ is said to be *satisfiable in* $\mathcal{S}$, if there is an assignment $h : Var \longrightarrow \mathrm{dom}(S)$ with $\mathcal{S} \models \Gamma[h]$. Γ is *satisfiable* if it is satisfiable in some structure. An assignment h to $\mathcal{S}$ is *complete* if h is a mapping onto $\mathrm{dom}(S)$. The *basic diagram* of $\mathcal{S}$ *under* complete assignment h is the subset of $\mathcal{L}_{basic}$ made true in $\mathcal{S}$ via h. The class of models of Γ is denoted by $MOD(\Gamma)$.

3.2 Environments

Let SEQ denote the collection of all finite sequences over the set $\mathcal{L}_{basic}$ of atomic formulas on vocabulary $\mathcal{L}$ and their negations. We define an *environment* to be any infinite sequence over $\mathcal{L}_{basic}$. Thus, for all $\sigma \in SEQ$, there is an environment e such that $\sigma = e|_n$ with $n = |\sigma|$. In this technical sense, SEQ then denotes the collection of all proper initial sequences of any environment. To consider consistent data-streams, we need to relate them to a structure. We do this in the next definition.

(1) DEFINITION: Let structure $\mathcal{S}$ and complete assignment h to $\mathcal{S}$ be given. An environment e is *for* $\mathcal{S}$ *via* h just in case $range(e) = \{\beta \in \mathcal{L}_{basic} \mid \mathcal{S} \models \beta[h]\}$. An

[2] By "infinite sequence" we shall always mean an ω-sequence, or a total function defined on N.

environment e is *for* $\mathcal{S}$ just in case e is an environment for $\mathcal{S}$ via some complete assignment.

In other words, an environment for a structure $\mathcal{S}$ via complete assignment h lists the basic diagram of $\mathcal{S}$ under h.

3.3 Basic components

The basic components of model-coordination are now to be introduced in detail. They are "agents," "interaction," and "success." We consider each in turn.

(2) DEFINITION: An agent in a model-coordination paradigm, or *basic agent*, is a pair $\langle \Psi , \mathbf{A} \rangle$, where Ψ is any mapping of *SEQ* into $\mathcal{L}_{basic}$, and $\mathbf{A}$ is a nonempty class of countable structures.

Faced with data $\sigma \in SEQ$, a basic agent $\langle \Psi , \mathbf{A} \rangle$ *believes* the *action* $\Psi(\sigma)$ if there is $\mathcal{S} \in \mathbf{A}$ such that $\mathcal{S} \models \Psi(\sigma)[h]$ for some (complete) assignment h to $\mathcal{S}$. If this is the case, we call action $\Psi(\sigma)$ *belief.* (In the terminology of Section 2, the pair $\mathcal{S}, h$ is called "knowledge.") For basic agent $\langle \Psi , \mathbf{A} \rangle$, the first component Ψ is called *communication function* or "ability". The second component $\mathbf{A}$ is called *background world.* We say that $\langle \Psi , \mathbf{A} \rangle$ is *based on* $\mathbf{A}$. We write Λ^b for the class of all basic agents and $\Lambda^b(\mathbf{A})$ for the class of all basic agents based on $\mathbf{A}$.

We now consider the kind of protocols, or "interaction," between two basic agents. For simplicity, we restrict attention to dynamics based on simultaneous moves, that is, the agents make decisions at the same time, and to pairwise communication only, that is, interactions involve just two agents. Our presentation may be generalized to a number of different protocols—examples are sequential moves and n-person communication, the choice only depending on what requires the specific scenario or application.

(3) DEFINITION: Let basic agents $\langle \Psi , \mathbf{A} \rangle$ and $\langle \Phi , \mathbf{B} \rangle$ be given.

1. The *interaction sequence* (or "play") of $\langle \Psi , \mathbf{A} \rangle$ and $\langle \Phi , \mathbf{B} \rangle$ is the infinite sequence
$$D_{\Psi,\Phi} = (\langle \overline{\Psi}_i , \overline{\Phi}_i \rangle \, : \, i \in N),$$
where $\overline{\Psi}_i$ is the ith move of Ψ and $\overline{\Phi}_i$ is the ith move of Φ, defined by induction as follows.
   i. $\overline{\Psi}_0 = \Psi(\emptyset)$ and $\overline{\Phi}_0 = \Phi(\emptyset)$.
   ii. $\overline{\Psi}_{n+1} = \Psi(\overline{\Phi}[n])$ and $\overline{\Phi}_{n+1} = \Phi(\overline{\Psi}[n])$.
2. The *response sequence*
$$\overline{\Psi} = (\overline{\Psi}_i \, : \, i \in N)$$
is the sequence of moves by basic agent Ψ in response to basic agent Φ, and the *response sequence*
$$\overline{\Phi} = (\overline{\Phi}_i \, : \, i \in N)$$
is the sequence of moves by Φ in response to Ψ.

3. Let $k \in N$ be given. The *interaction sequence* of $\langle \Psi, \mathbf{A} \rangle$ and $\langle \Phi, \mathbf{B} \rangle$ *starting at* k is the infinite sequence

$$^{k}D_{\Psi,\Phi} = (\langle\, _{k}|\overline{\Psi}_{i}\,, {}_{k}|\overline{\Phi}_{i}\,\rangle \;:\; i \in N),$$

where $_{k}|\overline{\Psi}_{i}$ is the ith element in $_{k}|\overline{\Psi}$ and $_{k}|\overline{\Phi}_{i}$ is the ith element in $_{k}|\overline{\Phi}$.

We then say that $_{k}|\overline{\Psi}_{i}$ is the ith move of Ψ starting at k and $_{k}|\overline{\Phi}_{i}$ is the ith move of Φ starting at k. Notice that $\overline{\Psi}$ is finite if and only if at any interaction step $i \in N$, $\Phi(\overline{\Psi}|_{i})$ or $\Psi(\overline{\Phi}|_{i})$ is undefined. If this is the case, it is immediate to verify that also $\overline{\Phi}$ is finite. Observe that $D_{\Psi,\Phi} = \langle\, \overline{\Psi}_{0}\,, \overline{\Phi}_{0}\,\rangle^{0}D_{\Psi,\Phi}$. Sometimes we write $R(\Psi, \Phi)$ for $\overline{\Psi}$ and $R(\Phi, \Psi)$ for $\overline{\Phi}$.

To coordinate, basic agents have to negotiate in the limit a 'complete and consistent' description of a structure in their own background world, eventually after a finite number of disagreements. The agents can restart their interaction finitely often, but after the last disagreement they must eventually coordinate. We now consider agents' "success."

(4) DEFINITION: Let basic agents $\langle \Psi, \mathbf{A} \rangle$ and $\langle \Phi, \mathbf{B} \rangle$ be given. We say that $\langle \Psi, \mathbf{A} \rangle$ *m-coordinates with* $\langle \Phi, \mathbf{B} \rangle$ (written: $\Psi \rightleftharpoons_{m} \Phi$) just in case for some $s, t \in N$:

- a $_{s}|\overline{\Psi}$ is an environment for some $\mathcal{A} \in \mathbf{A}$,
- b $_{t}|\overline{\Phi}$ is an environment for some $\mathcal{B} \in \mathbf{B}$, and
- c for all $n \in N$, $_{s}|\overline{\Psi}|_{n}$ is satisfiable in some $\mathcal{B}' \in \mathbf{B}$ and $_{t}|\overline{\Phi}|_{n}$ is satisfiable in some $\mathcal{A}' \in \mathbf{A}$.

Note that the definition depends on the interaction sequence $D_{\Psi,\Phi}$, and that structures $\mathcal{A}'$ and $\mathcal{B}'$ depend on n. In case of clause a, we say that Ψ *enumerates with* Φ $_{s}|\overline{\Psi}$; similarly, in case of clause b, we say that Φ *enumerates with* Ψ $_{t}|\overline{\Phi}$. So, to m-coordinate each basic agent outputs in the limit the basic diagram (under some complete assignment) of a structure in his own background world, and such basic diagram must be *finitely* consistent in some structure of the other agent's background world.

3.4 Two examples

At this point, one could ask why we require only finite consistency rather than full satisfiability (hence, isomorphism of the structures which the environments are for) in Definition (4)c. At every position of the interaction sequence, a basic agent knows only a finite part of the environment played by the other agent, so that he cannot know whether or not such environment is satisfiable in the actual structure he has in mind. This seems natural. A nice example where coordination in the sense of finite satisfiability is possible, but coordination in the sense of full satisfiability is not possible is now to be presented. We rely on the following definition of "exact-full" coordination.

(5) DEFINITION: Let basic agents $\langle \Psi, \mathbf{A} \rangle$ and $\langle \Phi, \mathbf{B} \rangle$ be given. We say that $\langle \Psi, \mathbf{A} \rangle$ *ef-coordinates with* $\langle \Phi, \mathbf{B} \rangle$ (written: $\Psi \rightleftharpoons_{ef} \Phi$) just in case for some $s, t \in N$:

a ${}_s|\overline{\Psi}$ is an environment for some $\mathcal{A} \in \mathbf{A}$,
b ${}_t|\overline{\Phi}$ is an environment for some $\mathcal{B} \in \mathbf{B}$, and
c ${}_s|\overline{\Psi}$ is satisfiable in some $\mathcal{B}' \in \mathbf{B}$ and ${}_t|\overline{\Phi}$ is satisfiable in some $\mathcal{A}' \in \mathbf{A}$.

Now comes the nice example.

(6) *Example:* Suppose that $\mathcal{L}$ is limited to a binary predicate symbol R. Let $\mathbf{A}$ be the class of all countable linear orders with maximum that interpret R, and let $\mathbf{B}$ be the class of all countable linear orders without maximum that interpret R. Then, (I) there are abilities Ψ, Φ such that $\langle \Psi, \mathbf{A} \rangle$ m-coordinates with $\langle \Phi, \mathbf{B} \rangle$, and (II) for all abilities Ψ, Φ, no basic agents $\langle \Psi, \mathbf{A} \rangle$ and $\langle \Phi, \mathbf{B} \rangle$ ef-coordinate. To prove (I) observe that any initial segment of any environment for a linearly ordered structure (with or without maximum) is satisfiable in any infinite linearly ordered structure, since it generates a substructure of the infinite linearly ordered structure and basic formulas are preserved in overstructures [Hod93, Thm. 2.4.1]. Define Ψ and Φ such that $R(\Psi, \Phi)$ is an environment for some infinite structure in $\mathbf{A}$ and $R(\Phi, \Psi)$ is an environment for some infinite structure in $\mathbf{B}$.[3] Then it is easy to see that for all $n \in N$, $\overline{\Psi}|_n$ is satisfiable in some structure in $\mathbf{B}$ and $\overline{\Phi}|_n$ is satisfiable in some structure in $\mathbf{A}$. Thus, $\langle \Psi, \mathbf{A} \rangle$ m-coordinates with $\langle \Phi, \mathbf{B} \rangle$. On the other hand, it is clear that (II) holds, since no environment for any structure in $\mathbf{A}$ is satisfiable in any structure in $\mathbf{B}$—the two classes $\mathbf{A}$ and $\mathbf{B}$ have no elementarily equivalent members in common.

However, not every m-negotiation game has an equilibrium. Before providing an example of such, we fix the following terminology.

(7) DEFINITION: Let nonempty sets $\Sigma(\mathbf{A})$ and $\Sigma(\mathbf{B})$ of basic agents based on, respectively, $\mathbf{A}$ and $\mathbf{B}$ be given. Let

$$\{D_{\Psi,\Phi}\} = \{D_{\Psi,\Phi} \mid \langle \Psi, \mathbf{A} \rangle \in \Sigma(\mathbf{A}), \langle \Phi, \mathbf{B} \rangle \in \Sigma(\mathbf{B})\}.$$

A *meaning negotiation* (in short: m-negotiation) *game* is a triple

$$\langle\, \Sigma(\mathbf{A}), \Sigma(\mathbf{B})\,, \{D_{\Psi,\Phi}\}\,\rangle.$$

An *equilibrium* of a m-negotiation game is a pair $(\mathcal{A}, \mathcal{B})$ of structures that satisfies conditions a–c of the definition above for some pair $\langle \Psi, \mathbf{A} \rangle \in \Sigma(\mathbf{A})$ and $\langle \Phi, \mathbf{B} \rangle \in \Sigma(\mathbf{B})$. A *partial equilibrium* (of a m-negotiation game) is a pair $(\mathcal{A}', \mathcal{B}')$ of structures that satisfies condition c of the definition.

The basic idea behind the notion of "equilibrium" is that an equilibrium should specify not only the basic agents' ability in the limit but also their final beliefs about the coordination history that occurred. The notion of partial equilibrium

[3]By "infinite structure" it is always meant a structure with infinite domain.

is less stringent; it specifies the agents' beliefs at *each* stage of the coordination process. Of course, an equilibrium and a partial equilibrium coincide as a special case. Also note that neither an equilibrium nor a partial equilibrium of a m-negotiation game is necessarily unique. An equilibrium does not depend on game stages, while a partial equilibrium does.

The next example shows that m-coordination is not always possible, that is, for some m-negotiation games an "equilibrium" does not exist.

(8) *Example:* Suppose that $\mathcal{L}$ is limited to unary predicate symbols H and T. [Read H as "Head" and T as "Tail".] Let $\theta \in \mathcal{L}_{sen}$ be of the form: $\forall x\,(Tx \leftrightarrow \neg\, Hx)$, $\vartheta \in \mathcal{L}_{sen}$ be of the form: $\forall x\, Tx$, and $\delta \in \mathcal{L}_{sen}$ be of the form: $\forall x\, Hx$. Let $\mathbf{A} = \mathit{MOD}(\{\theta, \vartheta\})$ and let $\mathbf{B} = \mathit{MOD}(\{\theta, \delta\})$. It is clear that the m-negotiation game $\langle\, \Lambda^b(\mathbf{A}), \Lambda^b(\mathbf{B})\,, \{D_{\Psi,\Phi}\}\,\rangle$ has no equilibrium, since no pair $\langle \Psi\,, \mathbf{A}\rangle$, $\langle \Phi\,, \mathbf{B}\rangle$ of basic agents can m-coordinate. [Recall that $\Lambda^b(\mathbf{A})$ denotes the class of all basic agents based on $\mathbf{A}$.]

The example concerns a game where players have diametrically opposed interests ($\mathbf{A} \cap \mathbf{B} = \emptyset$). It is called 'strictly competitive' by game theorists.

4 Competence of Basic Agents

A general question to ask about any paradigm of coordination concerns the existence of classes of agents such that the coordination competence of each agent in a class cannot be strictly improved. In this section, we exhibit a class of total, "self-centered" basic agents that has the required property for m-coordination. We rely on the following definition.

(9) Definition: We say that basic agent $\langle \Psi\,, \mathbf{A}\rangle$ is *self-centered* just in case for every $\Phi : \mathit{SEQ} \to \mathcal{L}_{basic}$, there is $\mathcal{A} \in \mathbf{A}$ such that $R(\Psi, \Phi)$ is an environment for $\mathcal{A}$.

In other words, a self-centered basic agent enumerates with some other basic agent an environment for some structure in the background world the agent is based on.

We now show that the negotiation competence of any total, self-centered basic agent cannot be improved by any total basic agent with the same background knowledge. We rely on the following sense of "meaning negotiation competence."

(10) Definition: The *m-negotiation competence* of a basic agent $\langle \Psi\,, \mathbf{A}\rangle$ is the set

$$m\text{-}\mathit{scope}(\Psi_{\mathbf{A}}) = \{\langle \Phi\,, \mathbf{B}\rangle \in \Lambda^b \mid \Psi \rightleftharpoons_m \Phi\}.$$

(11) Definition: Basic agents $\langle \Psi\,, \mathbf{A}\rangle$, $\langle \Phi\,, \mathbf{A}\rangle$ are *distinct* just in case $\Psi(\sigma) \neq \Phi(\sigma)$ for some $\sigma \in \mathit{SEQ}$.

(12) Definition: Let $\sigma \in \mathit{SEQ}$ and basic agent $\langle \Psi\,, \mathbf{A}\rangle$ be given. We say that $\langle \Psi\,, \mathbf{A}\rangle$ *starts with* σ just in case for all $\tau \in \mathit{SEQ}$ with $|\,\tau\,| < |\,\sigma\,|$, $\Psi(\tau) = (\sigma)_{|\,\tau\,|}$.

(13) THEOREM: For all distinct, self-centered basic agents $\langle \Psi , \mathbf{A} \rangle$, $\langle \Psi' , \mathbf{A} \rangle$, there is a basic agent such that $\langle \Psi , \mathbf{A} \rangle$ m-coordinates with and that $\langle \Psi' , \mathbf{A} \rangle$ does not.

We find an agent $\langle \Phi , \mathbf{A} \rangle$ such that, informally, behaves as follows. At step 0 Φ starts with σ. Then, if there is no previous move by any opponent basic agent to look at, Φ moves "safe". Otherwise, if an opponent basic agent "agrees" with basic agent $\langle \Psi , \mathbf{A} \rangle$ on σ, then Φ starts copying the opponent's moves from the first one. If the opponent basic agent "disagrees" with $\langle \Psi , \mathbf{A} \rangle$ on σ, then Φ breaks off coordination and starts producing inconsistent information.

Proof of Theorem (13): Let self-centered basic agents $\langle \Psi , \mathbf{A} \rangle$ and $\langle \Psi' , \mathbf{A} \rangle$ be distinct. Then there is $\sigma \in SEQ$ such that $\Psi(\sigma) \neq \Psi'(\sigma)$.

(14) Define ability $\Phi : SEQ \longrightarrow \mathcal{L}_{basic}$ such that:

1. Φ starts with σ;
2. for all $\tau \in SEQ$ with $|\tau| = |\sigma|$, $\Phi(\tau) = (v_0 \doteq v_0)$;
3. for all $\tau \in SEQ$ with $|\tau| > |\sigma|$, if $\tau_{|\sigma|} = \Psi(\sigma)$ then $\Phi(\tau) = \Psi(\tau|_{|\tau|-|\sigma|})$;
4. for all $\tau \in SEQ$ with $|\tau| > |\sigma|$, if $\tau_{|\sigma|} \neq \Psi(\sigma)$ then $\Phi(\tau) = \neg (v_0 \doteq v_0)$.

We claim that:

a $\langle \Psi , \mathbf{A} \rangle$ m-coordinates with $\langle \Phi , \mathbf{A} \rangle$.
b $\langle \Psi' , \mathbf{A} \rangle$ does not m-coordinate with $\langle \Phi , \mathbf{A} \rangle$.

(The proof of the claim may be found in [AM01], [Ago01, Prop. 2.(49)].) ∎

(15) COROLLARY: Let $\langle \Psi , \mathbf{A} \rangle$ be any total self-centered basic agent. Then there is no total basic agent $\langle \Psi' , \mathbf{A} \rangle$ such that $m\text{-}scope(\Psi_{\mathbf{A}}) \subset m\text{-}scope(\Psi'_{\mathbf{A}})$.

That is, the meaning negotiation competence of a total, self-centered basic agent cannot be strictly improved.

5 Computable Agents

How does our abstract paradigm provide the basis for a computational approach to the motivating problem of coordination by meaning negotiation? The focus of the present section is to provide a first answer to this question. Informally, our result says that there exists a computable agent who coordinates with every computable agent taken from an uniform class of computable agents. As a consequence of our result, we provide a procedure of pairwise coordination under reasonable assumptions.

To proceed formally, we narrow the interpretation of basic agents to computable objects. Computability thus enters our basic model in two related ways. First, the simplest way, the computability constraint is limited to basic agents' communication function. Second, the computability constraint is added to basic agents' background world. Crucial to the latter enterprise is finding a way to represent nonempty collections of structures in a computable manner. There are several possibilities for the definition of a "computable structure". We say that a structure $\mathcal{S}$ is *computable* just in case there is a computable, complete assignment h to $\mathcal{S}$ such that the basic diagram of $\mathcal{S}$ under h is a computable set.

From now on, we stipulate the convention that $\mathcal{L}_{form}$ is computable. Following [Mil99], we say that a first-order language is called *computable* (or "effectively presented") if the sets of its relation, function, and constant symbols, and the sets of variable and logical symbols are computable sets. Moreover, the functions that associate the relation and function symbols to their arity are computable.

(16) DEFINITION: A basic agent $\langle \Psi , \mathbf{A} \rangle$ is *computable* just in case Ψ is a computable function and $\mathbf{A}$ is a computable class of computable structures.

Now we fix in standard way a recursive isomorphism between N and *SEQ*.[4] (Recall that by convention above, *SEQ* is a countable and computable set of finite objects.) We also fix an acceptable indexing $\{\mathcal{W}_i^{bas} \,|\, i \in N\}$ of the *r.e.* subsets of $\mathcal{L}_{basic}$. Let $\{\varphi_i : SEQ \longrightarrow \mathcal{L}_{basic} \,|\, i \in N\}$ be an acceptable ordering of all (partial) computable communication functions from *SEQ* to $\mathcal{L}_{basic}$. So, every communication function Ψ is such that $\Psi = \varphi_i$ for some $i \in N$.

Let *STR* denote the class of all computable structures that interpret $\mathcal{L}$. Let $\{\mathbf{A}_i \subseteq STR \,|\, i \in N\}$ be an acceptable ordering of all computable classes of computable structures. Then $\{\langle \varphi_i , \mathbf{A}_i \rangle \,|\, i \in N\}$ is a recursive list of all computable basic agents.

5.1 Coordination of uniform classes

Let m_c denote the computable variant of the m-coordination paradigm, that is, the paradigm of m-coordination restricted to computable agents. Let computable basic agent $\langle \Phi , \mathbf{B} \rangle$ and nonempty collection Σ of computable basic agents be given. We say that $\langle \Phi , \mathbf{B} \rangle$ *m_c-coordinates with* Σ just in case $\langle \Phi , \mathbf{B} \rangle$ m-coordinates with every member of Σ. In this case, Σ is said to be *m_c-tractable.*

We recall that a collection Σ of computable basic agents is said to be *uniformly computable* just in case $\Sigma = \{\langle \varphi_{f(i)} , \mathbf{A}_{f(i)} \rangle \,|\, i \in N\}$ for some total computable function f from N to N.

(17) THEOREM: Every uniformly computable collection Σ of self-centered computable basic agents is $m_c$tractable.

Before proving the theorem, we provide a definition.

(18) DEFINITION: Let $\sigma \in SEQ$ and basic agent $\langle \Psi , \mathbf{A} \rangle$ be given. We define the *agent sequence* $\overline{\Psi(\sigma)} \in SEQ$ by induction on the length of σ as follows. $\overline{\Psi(\emptyset)} = \Psi(\emptyset)$. Suppose that $\overline{\Psi(\tau)}$ is defined for $\tau \in SEQ$, and that $\beta \in \mathcal{L}_{basic}$ is given. Then $\overline{\Psi(\tau\beta)} = \overline{\Psi(\tau)}\Psi(\tau\beta)$.

[4] In what follows, we assume elementary concepts of recursion theory as developed, for example, in [Rog67].

Notice that when $\overline{\Psi(\sigma)}$ is defined, its length is positive. We denote by $\overline{\Psi(\sigma)}^{-}$ the result of removing the last (rightmost) element from $\overline{\Psi(\sigma)}$.

Proof of Theorem (17): Let total computable function u from N to N be such that for all $i \in N$, basic agent $\varphi^i = \langle \varphi_{u(i)}, \mathbf{K}_{u(i)} \rangle$ is self-centered. Set $\Sigma = \{\langle \varphi_{u(i)}, \mathbf{A}_{u(i)} \rangle \mid i \in N\}$. Observe that Σ is uniformly computable. We define a basic agent $\langle \Phi, \mathbf{B} \rangle$ that m_c-coordinates with Σ as follows. Define $\mathbf{B} = \bigcup_{\varphi^i \in \Sigma} \mathbf{A}_{u(i)}$. Then, define Φ by induction as follows. Let $\Phi(\emptyset) = \varphi_{u(0)}(\emptyset)$. Suppose that $\Phi(\sigma)$ is defined on every $\sigma \in \mathit{SEQ}$ with $|\sigma| \leq n$. Given $\beta \in \mathcal{L}_{basic}$ and $\sigma \in \mathit{SEQ}$ with $|\sigma| = n$ such that $\mathit{range}(\sigma)$ is satisfiable in some $\mathcal{A} \in \mathbf{A}_{u(n)}$, let $i(\sigma) \in N$ be greatest with $i(\sigma) \leq n$ and $\mathit{range}(\overline{\varphi_{u(j)}(\overline{\Phi(\sigma)}^{-})}^{-})$ unsatisfiable in $\mathcal{A}$ for all $j < i(\sigma)$. Define $\Phi(\sigma\beta) = \varphi_{u(i(\sigma))}(\overline{\Phi(\sigma)})$. By the choice of u and the definition of $\mathbf{B}$, it follows that $\langle \Phi, \mathbf{B} \rangle$ is self-centered and computable.

To see that $\langle \Phi, \mathbf{B} \rangle$ m_c-coordinates with Σ, let $\varphi^n \in \Sigma$ be given. Since φ^n is self-centered, $R(\varphi^n, \Phi)$ is an environment for some $\mathcal{A}' \in \mathbf{A}_{u(n)}$. By the definition of Φ it follows that there is $m \leq n$ such that $R(\Phi, \varphi^m)$ is an environment for $\mathcal{A}'$. In particular, $R(\Phi, \varphi^n)$ is an environment for $\mathcal{A}'$. Then it is easy to verify that $\langle \Phi, \mathbf{B} \rangle$ m_c-coordinates with φ^n, hence Σ is m_c-tractable. ■

In other words, the coordination strategy adopted by Φ is the following. If there is no previous move by any opponent agent in Σ to look at, move "safe" by playing with the agent chosen according to total computable "function of choice" u at step 0. Otherwise, play with the "best choice agent" in Σ, who is playing consistently with his background word—the best agent be chosen according to u and the previous moves by other agents in Σ.

As a link to the game in Section 2, we note that for Σ be a class of "sources" ("users"), the proof of Theorem (17) exhibits a computable "seeker" who coordinates with every source in the class.

6 Related Work

Even if restricting the study of coordination to a first-order setting—as we did, the variety of interesting paradigms is clearly huge. A related discussion of model-coordination can be found in [AdJM00, Ago00].

Our approach is closed in spirit to the framework of the *language games*, namely, models of language change and language evolution in populations of communicating agents. Some reference are [Ste96, Bat00, Kir00]. Language games, however, focus much more on language creation and evoution than our models. Moreover, a formal comparison with language games would only be possible if we enriched our basic paradigm with an explicit system of "pay-offs," so as to explicitly relate agent's actions and beliefs. Other related models include semantic evaluation games (see for instance [HS97] for a survey), dialogue games for validity, and especially naming games—a kind of language games first introduced by Steels [Ste96]. As in our paradigm, in naming games "the meaning" play an important role. A speaker (the source) chooses a meaning and a form to express that meaning, and a hearer (the seeker) makes, based on the received form, a guess of what is meant. The hearer receives *direct* feedback on whether its guess is correct. A naming game is successful if the speaker's intention and the hearer's guess on it are the same, and fails otherwise. Our paradigm is similar to a naming game, as both serve to model systems for studying the emergence of conventional form-meanings associations. With at least one important difference: in our paradigm, the relation action-meaning has *induced* from the (hidden) knowledge made actual by an agent, so that the form to express meaning in our model is *indirect*.

7 Summary and Future Work

In this paper, we have advanced a formal paradigm of coordination and a computational variant thereof, wherein meaning negotiation plays a major role in the process of convergence to a common agreement. We have showed why the coordination competence of a special kind of "self-centered" agents cannot be strictly improved, and have exhibited a computable agent who coordinates with every member of an uniformly computable class of computable agents.

While we believe we have made progress on the coordination problem through meaning negotiation, we do not claim to have solved it in the general case. The paradigm can be modified in various ways—we mention two.

At the borderline of our approach is the problem of belief changes—how an agent should revise her beliefs upon learning new information. At present, belief revision is not made explicit in our model. On the other hand, we have seen that our paradigm considers an agent's beliefs to be represented by "background worlds." New paradigms of inductive meaning negotiation among 'revision-based' rather than 'knowledge-based' agents could then be advanced from our work.

Data available to the agents can be limited to just atomic formulas (no negations), or enriched to include universal or other kinds of formulas. Many alternative definitions of an "environment" are possible: incomplete, noisy, imperfect, recursive, multi-, etcetera. We chosen to consider "complete" environments in this paper as it seems the more simple solution. This is, however, a questionable representation of many natural environments. It is worth noting that "real" environments may suffer omissions, erroneous intrusions, or both omissions and intrusions. However, these important kinds of environments are outside the scope of this paper. They will form the focus of future work.

References

[AdJM00] A. Agostini, D. de Jongh, and F. Montagna. Coordination of 01 agents vs. coordination of worlds-based agents. Technical Report PP-2000-15, ILLC, Universiteit van Amsterdam, Amsterdam, NL, December 2000.

[Ago00] A. Agostini. Notes on formalizing coordination. In E. Lamma and P. Mello, editors, *AI*IA 99: Advanced in Artificial Intelligence*, pages 285–296. Springer-Verlag LNAI 1792, 2000.

[Ago01] A. Agostini. *Paradigms of Coordination and Solvability.* PhD thesis, Università di Siena, March 2001. Electronically available at `http://sra.itc.it/people/agostini/iphdt/idx.html`.

[AM01] A. Agostini and F. Montagna. Coordination and co-operation in the limit. Electronically available at `http://sra.itc.it/people/agostini/earchive.html`, October 2001.

[Bat00] J. Batali. The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In T. Briscoe, editor, *Linguistic Evolution through language acquisition: formal and computational models.* Cambridge University Press, Cambridge, 2000.

[Hod93] W. Hodges. *Model Theory.* Encyclopedia of mathematics and its applications, v. 42. Cambridge University Press, Cambridge, 1993.

[HS97] J. Hintikka and G. Sandu. Game-theoretical semantics. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 361–410. Elsvier Science B.V., Amsterdam, 1997.

[JFL+01] N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, C. Sierra, and M. Wooldridge. Automated Negotiation: Prospects, methods and challenges. *International Journal of Group Decision and Negotiation*, 10(2):199–215, 2001.

[JORS99] S. Jain, D. Osherson, J. Royer, and A. Sharma. *Systems That Learn - An Introduction to Learning Theory, 2nd edition.* The MIT Series in Learning, Development, and Conceptual Change, v. 22. The MIT Press, Cambridge, MA, 1999.

[Kel96] K. T. Kelly. *The Logic of Reliable Inquiry.* Oxford University Press, New York, NY, 1996.

[Kir00] S. Kirby. Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, J. Hurford, and M. Studdert-Kennedy, editors, *The Evolutionary Emergence of Language: Social function and the origins of linguistic form.* Cambridge University Press, Cambridge, 2000.

[Mil99] T. S. Millar. Pure Recursive Model Theory. In E. R. Griffor, editor, *Handbook of Computability Theory*, Studies in logic and the fundations of mathematics, v. 140, pages 507–532. Elsevier Science B.V., 1999.

[MO98] E. Martin and D. Osherson. *Elements of Scientific Inquiry.* The MIT Press, Cambridge, MA, 1998.

[OdJMW97] D. Osherson, D. de Jongh, E. Martin, and S. Weinstein. Formal Learning Theory. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 737–775. Elsevier Science B.V., 1997.

[Rog67] H. Rogers. *Theory of recursive functions and effective computability.* McGraw-Hill series in higher mathematics. McGraw-Hill, 1967.

[RZ94] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers.* The MIT Press, Cambridge, MA, 1994.

[Ste96] L. Steels. Self-organizing vocabularies. In C. Langton, editor, *Proceedings of the V Alife Conference*, Nara, Japan, 1996.

[Syc89] K. P. Sycara. Argumentation: Planning other agents' plans. In N. S. Sridharan, editor, *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, pages 517–523, San Mateo, CA, 1989. Morgan Kaufmann.

[WP00] M. Wooldridge and S. Parsons. Languages for negotiation. In W. Horn, editor, *Proceedings of the Fourteenth European Conference on Artificial Intelligence (ECAI-00)*, pages 393–397, Amsterdam, 2000. IOS Press.