Coordination: A model-theoretic perspective*

Alessandro Agostini — ITC-Irst Trento
Dick de Jongh — ILLC Amsterdam
Franco Montagna — Università di Siena

September 17, 2001

Abstract

A *learning to coordinate* paradigm was first introduced in *Formal Learning Theory* by [MO99] using the tools of recursion theory. In this paper, we advance and discuss a first-order paradigm of coordination—we call this paradigm of **model-coordination**. The paradigm is shown to extend Montagna and Osherson's paradigm of learning to coordinate, in the sense that Montagna and Osherson's binary players coordinate if and only if their first-order equivalent agents model-coordinate. An important difference between our paradigm and that proposed by [MO99] is that in our paradigm agents' preferences and beliefs can be modelled.

1 Introduction

Coordination is one of the most basic concepts in the foundations of the different sciences, be they physical, biological, behavioral or social, as it concerns the mutual interaction of possibly rational 'agents', some of them with individual beliefs, desires and intentions. In this paper, we focus on a family of coordination paradigms where the agents always move simultaneously and all relevant moves are made by the agents (agents' moving components are functions in the mathematical sense; no randomness ever intervenes). These paradigms are all based on pairwise communication, so that the kind of models we discuss is suitable for modeling group situations where communication is not with the whole group, as in an auction, but pairwise, as happens for instance in commercial transactions.

Our paradigms are games whose constitutive components we list as follows.

(1) **Components of our models of coordination:**

(a) a set of agents, or "players";

(b) a set of interaction-dynamics, or "plays", which provide information about the agents' interaction;

(c) a success criterion that stipulates when and under which conditions the agents can be said to coordinate.

*Electronic correspondence to: agostini@irst.itc.it, dickdj@wins.uva.nl, montagna@unisi.it.

The theory resulting from different paradigms of coordination as drawn by components in (1) is a bag of analytical tools designed to help us understand the phenomena that we observe when agents of the kind evoked in $(1)(a)$ interact.

Deductive or **introspective** theories that attempt to explain equilibrium play directly at the individual decision-making level impose very strong informational assumptions and so are widely recognized as having serious deficiencies. As a consequence, attention has shifted to purely inductive, or **behavioral** theories and methods, historically motivated by the work of evolutionary biologists. This approach involves exploring the *dynamics* of coordination actions. Of course, to do so requires the specification of the dynamic process describing the play of the agents concerned. Now, it is a matter of taste if the middle ground between these extremes is more interesting than either extreme. But, as far as we know, only isolated forays have so far been made from the standpoint of *formal learning theory* into this borderline area. The work on learning to coordinate choices by [MO99] is a recent one, and at last provides a recursion-theoretic approach. This article is a further investigation into this area but from a different, though related, perspective.

Specifically, we develop a new paradigm of coordination richer in structure than that considered by Montagna and Osherson—we call our paradigm: **model-coordination**. The basic notions of an agent, dynamics and coordination success we present extend the work by [MO99] to a first-order framework. These notions add a new perspective on the problem of equilibrium selection in pairwise coordination. Some emphasis is given to the concept of an 'agent' as composed by both communicative ('behavioral') capability and beliefs, and to the dynamic processes by means of which equilibrium is eventually achieved through iterated, interactive mechanisms based on some introspective, possibly rational individual choice. This allows us to show how agents' preferences and beliefs can influence coordination in a way not given in the work of Montagna and Osherson. An important difference between our paradigm and that proposed by [MO99] is that in our paradigm agents' preferences and beliefs can be modelled. We show how this is done, and also outline why the model of Montagna and Osherson could not handle beliefs by demonstrating that pairwise coordination between "players" in the sense of Montagna and Osherson arises as a special case of coordination between belief-based agents in our sense. This is indeed the main result of this paper.

Our aim is to reduce model-coordination to binary coordination in the sense of [MO99]. We begin in the next section by presenting a paradigm of basic coordination which concerns us. The paradigm is equivalent to the learning to coordinate paradigm by Montagna and Osherson, and it serves us as a base for theory development. In Section 3 a model-theoretic paradigm (more accurately, a family of paradigms) is introduced and defined within a first-order framework. Some comments on the basic components of the paradigm and a brief comparison with a slightly different paradigm of coordination (Section 4) are also given. Section 5 then provides our main result. We end the paper in Section 6 with a short conclusion.

2 The 01-Coordination Paradigm

"Sam and Sally like to meet daily in the park, pretending each time that it's yet another chance encounter, walking side by side in shy silence. Each shows up punctually at either noon or 6:00 p.m. hoping the other will have made the same choice. The shifting constraints on their schedules, however, make it hard to predict who will select which time of arrival, and both suffer disappointment when there is mismatch. So both Sam and Sally set about trying to predict the other's choices, desiring to act in concert. Their predictions are based on no more than the history of earlier events...." [MO99, p. 363]

A paradigm of "learning to coordinate choices" is introduced into *Formal Learning Theory* (see for instance [JORS99] for a survey) by Montagna and Osherson using the tools of recursion theory. [Kel96] advanced a similar paradigm without a formal development. In terms of the game of Sam and Sally, the paradigm is a common interest, repeated game. Two agents ("players") want to coordinate by repeatedly showing each other one of two possible behaviors. The problem of coordination the players are faced with follows from the shifting constraints on their behaviors. Each player tries to predict the other's behavior, and their predictions are based on no more than the history of earlier events. One player 'learns' the other's behavior just in case his or her own behavior matches the other's forever after. To keep matters simple, Montagna and Osherson's players face the same two options on each trial, conventionally denoted by 0, 1. A player is therefore a "bit agent", namely, any function from the set of all finite binary sequences into $\{0, 1\}$, where any such sequence is conceived as the history of moves of an opposing player. We call the kind of coordination between bit agents: **01-coordination**. We shall now specify the paradigm precisely. Our formal presentation relies on the following notation, some of which will also be used in the rest of this paper.

2.1 Notation

We denote the set $\{0, 1, 2, ...\}$ of natural numbers by N. Let η be an infinite sequence over $\{0, 1\}$.[1] For $i \in N$, we write $\eta|_i$ for the proper initial sequence of length i in η. We write $|\sigma|$ for the length of a finite sequence, $\emptyset$ for the finite sequence of length zero, σ_i or also $(\sigma)_i$ for the ith element of σ, $0 \leq i < |\sigma|$, $last(\sigma)$ for the last element in σ, and σ^- for the finite sequence obtained from σ by dropping its last (rightmost) element, if $\sigma \neq \emptyset$; σ^- is $\emptyset$, otherwise. Thus, $|\sigma^-| = |\sigma| - 1$, $last(\sigma) = \sigma_{|\sigma^-|}$ and $\sigma = \sigma^- last(\sigma)$. Concatenation of sequences is indicated by juxtaposition, or also by "$\frown$" when juxtaposition might result in unclear expressions, and we won't distinguish notationally between an element and the corresponding unit sequence. Thus $\alpha\tau$ ($\alpha^\frown\tau$) denotes the sequence with first element α and tail τ. We write $\sigma \sqsubseteq \tau$ if σ is a prefix of τ, that is, $\tau = \sigma\eta$ for some η. In this case, we say that τ **extends** σ. The set of elements in any sequence τ is denoted by $range(\tau)$.

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

2.2 Components

We now are ready to formalize the components listed in (1).

2.2.1 Agents.

Let *BISEQ* denote the collection of all finite sequences over $\{0, 1\}$.

(2) DEFINITION: A 01-**agent** (or "bit agent") is any mapping from *BISEQ* to $\{0, 1\}$.

01-agents can be partial or total, computable (in a precise sense given by coding *BISEQ* and $\{0, 1\}$) or uncomputable. We denote the set of all 01-agents by Λ^1_0. In particular, we recall from standard recursion-theoretic terminology that a set $\Sigma \subseteq \Lambda^1_0$ of partial computable 01-agents is called **uniformly computable** just in case there is a total recursive function f such that $\Sigma = \{\varphi_{f(i)} \mid i \in N\}$. Note that *BISEQ* includes all the finite proper segments arising from any 01-agent's input. This feature is fundamental to define the kind of pairwise interaction between 01-agents we are interested in. We do it in the next paragraph.

2.2.2 Dynamics.

Let 01-agents Ψ and Φ be given. The **interaction sequence** (or "play") $D_{\Psi,\Phi}$ of Ψ and Φ is built up as follows. Bit agents Ψ, Φ simultaneously apply to the empty sequence, yielding the first **moves** (positions, steps) $\overline{\Psi}_0 = \Psi(\emptyset)$ and $\overline{\Phi}_0 = \Phi(\emptyset)$. Then 01-agent Ψ applies to the sequence containing $\overline{\Phi}_0$ as its only element, yielding the second move by Ψ, namely, $\overline{\Psi}_1 = \Psi(\overline{\Phi}_0)$. Similarly, Φ applies to the sequence containing $\overline{\Psi}_0$ as its only element, yielding $\overline{\Phi}_1 = \Phi(\overline{\Psi}_0)$. In general, the nth move of Ψ is the value of Ψ on the sequence of all previous moves by Φ, and the nth move of Φ is the value of Φ on the sequence of all previous moves by Ψ. In the limit, the interaction of Ψ and Φ generates the infinite interaction sequence of pairs:

$$D_{\Psi,\Phi} = (\langle \overline{\Psi}_i , \overline{\Phi}_i \rangle \, : \, i \in N),$$

where $\overline{\Psi}_i$ is the ith move of bit agent Ψ and $\overline{\Phi}_i$ is the ith move of bit agent Φ. The **response sequence**:

$$R(\Psi, \Phi) = (\overline{\Psi}_i \, : \, i \in N)$$

is the (finite or infinite) sequence of moves by Ψ in response to Φ, and the **response sequence**:

$$R(\Phi, \Psi) = (\overline{\Phi}_i \, : \, i \in N)$$

is the (finite or infinite) sequence of moves by Φ in response to Ψ. Notice that $R(\Psi, \Phi)$ is finite iff at some 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 $R(\Phi, \Psi)$ is finite also. We are now ready to define success within the 01-coordination paradigm.

2.2.3 Success.

Next comes an instance of a success criterion (1)(c) that stipulates when and under which conditions the agents can be said to coordinate.

(3) DEFINITION: We say that bit agent Ψ **01-coordinates** with collection Σ of bit agents just in case for response sequences $R(\Psi, \Phi)$ and $R(\Phi, \Psi)$, the following conditions hold for each $\Phi \in \Sigma$:

(a) $R(\Psi, \Phi)$ and $R(\Phi, \Psi)$ are infinite sequences; and
(b) for cofinitely many $n \in N$, $R(\Psi, \Phi)_n = R(\Phi, \Psi)_n$.[2]

If some 01-agent 01-coordinates with Σ, then Σ is called 01-**coordinable**, and 01-**uncoordinable** otherwise.

The following example should suffice to explain the previous definition of the basic components of a 01-coordination paradigm.

(4) *Example:* Let bit agents Ψ and Φ be given.

(a) Suppose that Ψ and Φ are **rigid**, in the sense that both make their decisions regardless of the other agent's (last) move. In particular, any "constant bit agent", that is, either a mapping of *BISEQ* uniformly to 0 or a mapping of *BISEQ* uniformly to 1, is rigid. Now suppose that Ψ's response to Φ is always 0 and that Φ's response to Ψ is always 1. Then Ψ and Φ cannot 01-coordinate.

(b) Suppose that Ψ and Φ are **flexible**, in the sense that both are willing to imitate the other agent's behavior. This means that at any step of the play, Ψ is ready to change his "plan of action" by looking at the last move of Φ and then copying it in response to her last move, and vice versa. If we add the assumption that Ψ starts moving 0 and Φ starts moving 1 (this condition is indeed consistent with the assumption of flexibility, because there is no last move at step zero), it follows that Ψ and Φ never 01-coordinate. In fact, their response sequences are both infinite, namely, 010101... and 101010..., but these sequences never coincide, even after a finite number of mismatchings. However, if both Ψ and Φ start moving 0 (or 1), it is obvious to conclude that Ψ 01-coordinates with Φ in one step.

The next subsection provides a glimpse of 01-coordinable and 01-uncoordinable classes of bit agents.

2.3 Coordinable, uncoordinable classes

The following proposition exhibits a 01-coordinable collection of bit agents.

(5) PROPOSITION: [MO99] Every uniformly computable collection of total computable bit agents is 01-coordinable.

[2] By "cofinitely many" we shall always mean "for all $n \in N$ but a finite set".

Proof: We follow [MO99]. For all $\sigma \in BISEQ$ and all 01-agents Ψ, we first define $\overline{\Psi(\sigma)} \in BISEQ$ by induction on the length of σ as follows. *Base*: $\overline{\Psi(\emptyset)} = \Psi(\emptyset)$. Suppose that $\overline{\Psi(\tau)}$ is defined for $\tau \in BISEQ$ and let $b \in \{0, 1\}$ be given. Then, define $\overline{\Psi(\tau b)} = \overline{\Psi(\tau)}\Psi(\tau b)$. Let total recursive function $f : N \longrightarrow N$ be such that $\varphi_{f(i)}$ is total for all $i \in N$. We define by induction a 01-agent Φ that 01-coordinates with uniformly computable $\Sigma = \{\varphi_{f(i)} \mid i \in N\}$. *Base*: $\overline{\Phi(\emptyset)} = \varphi_{f(0)}(\emptyset)$. Suppose that $\overline{\Phi(\sigma)}$ is defined for $\sigma \in BISEQ$ with $|\sigma| \leq n$. Given $\tau \in BISEQ$ with $|\tau| = n$, let $i(\sigma) \in N$ be greatest with $i(\sigma) \leq n$, and $\overline{\overline{\varphi_{f(j)}(\overline{\Phi(\sigma)}^{-})}^{-}} \neq \sigma$ for all $j < i(\sigma)$. Then define $\Phi(\sigma b) = \varphi_{f(i(\sigma))}(\overline{\Phi(\sigma)})$. Clearly, Φ is total computable. Let now $\mathcal{X} \in \Sigma$ be given, and suppose $\mathcal{X} = \varphi_{f(n)}$. Then, for some $m \leq n$, there are cofinitely many $k \in N$ such that $R(\varphi_{f(n)}, \Phi)_k = R(\varphi_{f(m)}, \Phi)_k = R(\Phi, \varphi_{f(m)})_k$. Hence, Φ 01-coordinates with each member of Σ, that is, Σ is 01-coordinable. ■

The next proposition shows that restricting bit agents to be total reduces the possibilities of coordination.

(6) PROPOSITION: [MO99] There is a 01-coordinable collection of total bit agents that no total bit agent 01-coordinates with.

For the proof of the proposition see [MO99, Cor. 12]. In the next section, we motivate the need of a new paradigm of coordination, richer than 01-coordination, so to be able to encompass preferences and beliefs in the model.

2.4 A criticism

As a model of coordination, the 01-coordination paradigm is limited for at least two reasons. First, the behavior of the 01-agents is represented by bits of information, so that only two possible choices are admitted for each agent at any time. For greater realism, it would be better for, say, Sam and Sally to show up in the park more often than at noon and 6:00 p.m., even if the problem of coordination they face presumably becomes more complicated. This greater realism is possible if we represent the agents' action space by a collection of (finitary) first-order formulas. Second, the 01-coordination paradigm captures neither the agent's "beliefs" nor the agent's "preferences". For example, the story of Sam and Sally that "like to meet daily in the park" cannot be formalized within the paradigm. Suppose that Sam prefers to meet Sally in the park on Monday and Friday, but at that famous pub on Thursday, because live jazz music is played there on Thursday and he is a jazz music lover. (Sally continues to prefer the park every day.) Is then coordination possible between Sam and Sally? The question is not even expressible in the 01-coordination paradigm. Even if both choose identical meeting times, it is quite sure that they won't ever coordinate, because of their different, "uncoordinated" preferences. More generally, the 01-coordination paradigm does not fully capture situations where the players have different preferences. The paradigm is purely behavioral, the agents' predictions being based on no more than the history of earlier moves. However, an agent can use some background knowledge to act.

These and similar remarks motivate us in extending the paradigm to a first-order setting. For doing this, we will henceforth identify an "agent" with a pair of objects, where the first element of the pair models the agent's actions and the second element of the pair models the agent's preferences and beliefs.[3] A paradigm concerning agents of such form will be called of **model-coordination**, for short: m-coordination. Despite its greater expressiveness, however, model-coordination remains a crude representation of real coordination; "real" agents' actions are rarely carried out by literals of a first-order language, and preferences are far from being captured by arbitrary collections of first-order structures.

With these remarks in mind, we conclude our *résumé* of the 01-coordination paradigm, which leaves now room open to present the basic elements of a paradigm of model-theoretic coordination. Coordination is "model-theoretic" in the sense that should become clearer before the end of the next section.

3 Model-coordination

3.1 An intuitive picture

The following intuitive picture of a 2-agent infinitely repeated game is intended to help the reader interpret some of the abstract concepts described later.

(7) *Example:* We imagine two "knowledge-based" agents, say **A**lfonso and **B**arbara, whose preferences and beliefs are represented by two nonempty classes of first-order structures $\mathbf{A}$ and $\mathbf{B}$ that interpret a common, countable vocabulary $\mathcal{L}$. To coordinate, **A**lfonso and **B**arbara communicate with each other in order to respectively end up in the limit with a consistent description of a structure, $\mathcal{A} \in \mathbf{A}$ and $\mathcal{B} \in \mathbf{B}$, such that $\mathcal{A}$ is sufficiently close to $\mathbf{B}$ and $\mathcal{B}$ is sufficiently close to $\mathbf{A}$. We may think that $\mathcal{A}$ is sufficiently close to $\mathbf{B}$ if every finite set of literals true in $\mathcal{A}$ is satisfiable in some $\mathcal{B}' \in \mathbf{B}$.[4] Thus, we expect that the more **A**lfonso is like **B**arbara, that is, **A**lfonso and **B**arbara have similar preferences and beliefs, the better chance **A**lfonso and **B**arbara have to coordinate. To dramatize, let us suppose that each agent does not know the preferences of the other, and that the agents were never before in a similar situation, so they cannot rely on past experience to solve their coordination problem. **A**lfonso and **B**arbara's decisions have a strategic component, since **A**lfonso and **B**arbara take into account their own preferences and expectations as well as the *other* agent's behavior. Since strategic interactions are best modeled by game-theoretic approaches, we imagine the agents' coordination problem as a form of 2-player game. To start the game, **A**lfonso is conceived as choosing one member from $\mathbf{A}$ to be his "actual world", namely, the structure that models **A**lfonso's initial preferences. **A**lfonso's choice is initially unknown to **B**arbara. **A**lfonso then provides a "clue" (**A**lfonso's action) about his chosen world. At the same time (but we can imagine a similar paradigm where agents make their choices in sequence), **B**arbara makes her choice as well,

[3] For simplicity, from now to the end of this paper, we shall always consider the two terms "preference" and "belief" as synonymous.

[4] A slightly different version of the paradigm uses $\mathcal{B}$ in place of $\mathcal{B}'$; see [Ago01, Def. 2.(35)].

and provides **A**lfonso with a clue about her actual world. We can assume that **A**lfonso and **B**arbara are allowed to change their actual world at each step of the game, provided that they remain coherent with the actions shown beforehand. If **A**lfonso realizes that his preferences are not close enough to **B**arbara's actions, he can change it by choosing a different actual world, and this holds for **B**arbara as well. Each player may provide "bad clues" in principle, for example to inform the other player of the desire of starting again. Consequently, the players are allowed to start the game again from the beginning whenever some disagreement occurs. This may happen if a player's action is inconsistent with the set of preferences shown by the opponent since then. Of course, to reach coordination disagreement should happen only finitely often. **A**lfonso's clues constitute the data upon which **B**arbara will base her hypotheses on **A**lfonso's actual world, that eventually become themselves a clue for **A**lfonso about **B**arbara's world. And so on. Each time **A**lfonso provides a new clue, **B**arbara may produce a new hypothesis, and a new clue for **A**lfonso as well. **A**lfonso and **B**arbara succeed in their coordination game if their own actual world becomes in the limit sufficiently close to the actual world of the other agent, in the sense explained above.

The example motivates us in looking for a paradigm that captures as many relevant aspects as possible of the kind of game played by **A**lfonso and **B**arbara. Some concepts like "agents, "clues", and "success" figure in the foregoing picture of the **A**lfonso and **B**arbara coordination game. We formalize them in the next section by moving through the components listed in (1).

3.2 Components

We fix a first-order framework. For this purpose, we write $\mathcal{L}_{form}$ to denote the set of first-order **formulas** built up from a vocabulary $\mathcal{L}$ consisting of predicates (equality symbol is: $\doteq$) and function symbols of various arities, along with constants symbols, together with a countably infinite set $Var = \{x_i \mid i \in N\}$ of variables. To avoid ambiguities, we let formulas contain parentheses as auxiliary signs. We use $\mathcal{L}_{basic}$ and $\mathcal{L}_{sen}$ to denote, respectively, the set of **basic formulas** (that is, the atomic formulas and their negations) and the set of **sentences** of $\mathcal{L}_{form}$. We are particularly interested in the collection of all the finite sequences over $\mathcal{L}_{basic}$. We denote such collection by SEQ.

To make the paper self-contained, we recall some standard semantic notions. Let $\mathcal{S}$ be a structure that interprets $\mathcal{L}$. We denote the domain of $\mathcal{S}$ by $\mathrm{dom}(S)$. 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. In particular, a finite or infinite sequence τ on $\mathcal{L}_{form}$ is satisfiable if $range(\tau)$ is satisfiable. 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 a 'diagram' in the sense of Abraham Robinson, that is, the subset of $\mathcal{L}_{basic}$ made true in $\mathcal{S}$ via h.

3.2.1 Agents.

(8) DEFINITION: An agent in an m-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 structures.

Intuitively, faced with data-stream $\sigma \in SEQ$, a basic agent $\langle \Psi, \mathbf{A} \rangle$ believes **action** $\Psi(\sigma)$. Since class $\mathbf{A}$ is interpreted as representing $\langle \Psi, \mathbf{A} \rangle$'s beliefs, this means that $\Psi(\sigma)$ is expected to be true in some structure in $\mathbf{A}$.[5] 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 $\langle \Psi, \mathbf{A} \rangle$ (or also Ψ) is **based on A**, Ψ is **of** $\langle \Psi, \mathbf{A} \rangle$, and $\langle \Psi, \mathbf{A} \rangle$ **has** Ψ. We write $\mathbf{A}_\Psi$ for $\mathbf{A}$ and $\Psi_\mathbf{A}$ for Ψ (or also $\langle \Psi, \mathbf{A} \rangle$) just in case Ψ is based on $\mathbf{A}$. We refer to $\mathcal{L}$ as the agent's vocabulary, and to $\mathcal{L}_{basic}$ as the agent's language. It is because of the basic language $\mathcal{L}_{basic}$ that agents in the m-coordination paradigm are termed: "basic". A basic agent may have a partial, total, computable or noncomputable communication ability.

3.2.2 Dynamics.

Before returning to technical issues, we indulge in a general remark on dynamics. The paradigm of model-coordination is designed to examine the logic of long-term interaction. It captures the idea that an agent will take into account the effect of his current behavior on the other agent's future behavior, and aims to explain phenomena like cooperation and threats. In the following, we focus attention to dynamics based on simultaneous moves, that is, the agents make decisions at the same time, and to pairwise interaction. We note, however, that these restrictions may be removed without affecting the main result of this paper. We refer the reader to [Ago01] for a discussion of related paradigms.

We now consider formally the interaction between two basic agents. 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. Thus, $\zeta[k] = \zeta|_{k+1}$ (cf. section 2.1), and ${}_0|\zeta = \emptyset$ if $|\zeta| = 1$.

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

(a) 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 $\overline{\Phi}_0 = \Phi(\emptyset)$.

[5] This is not a strict requirement, however, but the underlying intuition should help the reader in understanding the criterion of success of any model-coordination paradigm. By the same interpretation, we require $\mathbf{A}$ to be nonempty, so that agents are assumed to believe something. (It is not too strong a requirement to ask agents to believe $x_0 \doteq x_0$!)

ii. $\overline{\Psi}_{n+1} = \Psi(\overline{\Phi}[n])$ and $\overline{\Phi}_{n+1} = \Phi(\overline{\Psi}[n])$.

(*b*) 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. Observe that $D_{\Psi,\Phi} = \langle\, \overline{\Psi}_0 \,, \overline{\Phi}_0 \,\rangle {}^0D_{\Psi,\Phi}$.

(10) *Remark:* The definition of interaction sequence $^{k}D_{\Psi,\Phi}$ depends only on the agents' abilities Ψ and Φ; no background worlds are involved. We shall see later in this section that basic agents' background worlds are relevant to determine the criterion of coordination success.

Similarly to the agents' response sequences within the 01-coordination paradigm (see subsection 2.2), the **response sequence**

$$R(\Psi, \Phi) = (\overline{\Psi}_i \;:\; i \in N)$$

is the (finite or infinite) sequence of moves by basic agent Ψ in response to basic agent Φ, and the **response sequence**

$$R(\Phi, \Psi) = (\overline{\Phi}_i \;:\; i \in N)$$

is the sequence of moves by Φ in response to Ψ. Again, notice that $R(\Psi, \Phi)$ is finite iff at any interaction step $i \in N$, $\Psi(\overline{\Psi}|_i)$ or $\Psi(\overline{\Phi}|_i)$ is undefined. If this is the case, it is immediate to verify that $R(\Phi, \Psi)$ is finite. Moreover, note that a response sequence is an environment. Consistently with the notation adopted on infinite sequences within the 01-coordination paradigm, ${}_k|R(\Psi, \Phi)|_n$ denotes the finite initial sequence in ${}_k|R(\Psi, \Phi)$ of length n, and ${}_k|R(\Psi, \Phi)_n$, or also $({}_k|R(\Psi, \Phi))_n$ denotes the nth element of ${}_k|R(\Psi, \Phi)$. So, ${}_k|R(\Psi, \Phi)|_{n+1} = {}_k|R(\Psi, \Phi)|_n \overline{\Psi}_n$ and $R(\Psi, \Phi) = \overline{\Psi}_0 {}^\frown {}_0|R(\Psi, \Phi)$.

3.2.3 Success.

To coordinate basic agents have to stabilize in the limit to a complete description of a structure in their own background world, eventually after a finite number of failures, or disagreements. The description that a basic agent produces in the limit is a kind of "environment", about which the other basic agent is concerned. An **environment** is then any infinite sequence over $\mathcal{L}_{basic}$. As a consequence, 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 produced by a basic agents' interaction, we need to relate an environment to a structure. We rely on the next definition.

(11) 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 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. Finite initial segments of an environment for a structure thus recapitulate the *consistent* information available to a basic agent about the underlying structure of evidence—we called it "actual world", at a certain time of observation.

(12) *Example:* Suppose that $\mathcal{L}$ is limited to a binary predicate symbol R plus equality symbol $\doteq$. Let $\langle N, < \rangle$ be the standard structure of natural numbers ordered according to strict linear order $<$. Let $h : Var \longrightarrow \mathrm{dom}(N)$ be defined such that for all $i \in N$, $h(x_i) = i$. Then one environment for $\langle N, < \rangle$ via h is:

$x_0 \doteq x_0 \ \neg x_0 R x_0 \ \ldots$

$\ldots \ x_1 \doteq x_1 \ \neg x_1 R x_1 \ x_0 R x_1 \ \neg x_1 R x_0 \ \ldots$

$\ldots \ x_n \doteq x_n \ \neg x_n R x_n \ x_0 R x_n \ \ldots \ x_{n-1} R x_n \ \neg x_n R x_0 \ \neg x_n R x_1 \ \ldots \neg x_n R x_{n-1} \ \ldots$

Observe:

(13) LEMMA: If some environment is for both structures $\mathcal{S}$ and $\mathcal{T}$, then $\mathcal{S}$ and $\mathcal{T}$ are isomorphic.

(14) LEMMA: All isomorphic pairs $\mathcal{S}$ and $\mathcal{T}$ of structures have an identical set of environments.

The proofs of the lemmas above are an immediate consequence of [Kei77, Prop. 3.2(i)]; the proof of Lemma (13) is also in [OW86, Lem. 3.1A].

According to the intuitive picture of model-coordination, the agents can restart their interaction finitely often, but after the last disagreement they must eventually coordinate. Let us now present all this formally.

(15) 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 this basic diagram must be finitely consistent with some structure of the other agent's background world—generally different from the structure described by him in the limit.

3.3 Comments on Components

An interpretation of the paradigm of m-coordination (see also [Ago01]) is that it is a model of an event occurring repeatedly, as in Example (7). So, each agent can form his expectation of the other agent's behavior on the basis of his (innate, previously experienced) knowledge as well as on the basis of information coming directly from the other agent's behavior at the time the event occurs. Neither a specific knowledge of the event nor common knowledge or agents' "rationality" is assumed. The agents choose their actions simultaneously and independently, and have the same language vocabulary as the unique common "communication tool".

From the definition of the interaction sequences it is clear that the paradigm of m-coordination is a model of a limiting process. The fact that the limit is taken to be finite or infinite, however, depends on the use of the model. Here we add the comment that the choice of a finite or infinite limit in the paradigm under analysis is also depending on the background world each agent concerned is based on. A quick look back to Definition (15) should suffice to realize that no finite interaction sequence may guarantee success if at least one of the two basic agents is based on a background world with only infinite structures; there is no finite environment for an infinite structure. Nevertheless, it is important to stress that *finite* interaction sequences in a m-coordination paradigm are possible and even necessary to model particular phenomena.

The agents can have a common interest to coordinate. In this sense, one game theorist would say that the paradigm is a *pure* coordination game. In fact, the paradigm is a (finitely, infinitely) repeated game of pure coordination with imperfect information. A model of **imperfect information** allows an agent, when taking an action, to have only partial information about the actions taken previously. Thus, the m-coordination paradigm is a model of imperfect information, since an action $\Psi(\sigma)$ by basic agent $\langle \Psi\, , \mathbf{A} \rangle$ on any $\sigma \in \mathit{SEQ}$ is eventually true in some structure $\mathcal{A} \in \mathbf{A}$ which is generally *unknown* to the other basic agent. We note that, on the other hand, the definition of response sequences makes $\Psi(\sigma)$ to be part of the input sequence of any basic agent Φ for all σ representing an initial segment of Φ's response sequence. Consequently, if we consider basic agents' communication functions only, then the information the agents interchange is "perfect", in the sense that each agent is perfectly informed about all actions executed by the other agent at each stage of the game.[6]

About the assumption that the agents act simultaneously, we observe that this does not necessarily mean that agents' actions are taken at the same point of time, even if this is our basic interpretation of the paradigm. Suppose, for example, that

[6]We refer to [OR94, Part II, Part III] for background on games with perfect and imperfect information, respectively.

Alfonso and **B**arbara of Example (7) are at different geographical locations, in front of terminals, trying to communicate through the Internet. Then, the usual net-delay forces **A**lfonso and **B**arbara' acts to be asynchronous, but their actions are still interpreted as simultaneous in our model. In the m-coordination paradigm it is only important that the agents make decisions independently at any stage of the coordination process, no agent being informed of the choice of the other agent prior to making his own decision.

4 Full satisfiability of beliefs

We require only finite consistency rather than full satisfiability (hence, isomorphism of the structures which the environments are for) in Definition (15)(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.

(16) 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.

(17) LEMMA: 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, the following conditions hold.

(a) There are abilities Ψ, Φ such that $\langle \Psi, \mathbf{A} \rangle$ m-coordinates with $\langle \Phi, \mathbf{B} \rangle$.

(b) For no abilities Ψ, Φ, any basic agents $\langle \Psi, \mathbf{A} \rangle$ and $\langle \Phi, \mathbf{B} \rangle$ ef-coordinate.

Proof: To prove (a), 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 superstructures; see [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 $\mathbf{B}$.[7] Then it is easy to see that for all $n \in N$, $\overline{\Psi}|_n$ is satisfiable in some structure in

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

$\mathbf{B}$ and $\overline{\Phi}|_n$ is satisfiable in some structure in $\mathbf{A}$. On the other hand, it is clear that (b) 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. ■

5 01-Coordination as a special case

In this section, we prove our main result. Model-coordination will be showed to extend 01-coordination, in the sense that two bit agents 01-coordinate if and only if some "first-order equivalent" agents m-coordinate. Our aim is to reduce the paradigm of model-coordination to binary coordination in the sense of Section 2.

For this purpose, we fix $\mathcal{L} = \{\dot{=}, U, s, \overline{0}\}$.[8] $\overline{0}$ and s denote, respectively, 0 and the successor function on the set of natural numbers. Thus, the term that results from $n \in N$ applications of s to 0 is denoted by $\overline{n}$; U denotes a set of natural numbers. This is also the place to find some new terminology and notation. Given a pair of sets A, B, we write $A \triangle B$ for the symmetric difference of A and B, i.e., the set $(A - B) \cup (B - A)$. If $A \triangle B$ is finite, we say that A is a **finite variant of** B and A and B are called **finite variants**. Observe that $A \triangle B$ is finite iff either $A = B$ or A and B differ by a finite set. For $U \subseteq N$, $\mathcal{N}_U$ denotes the structure $\langle N, U, s, 0\rangle$ that interprets $\mathcal{L}$. In other words, $\mathcal{N}_U$ is the usual model $\mathcal{N} = \langle N, s, 0\rangle$ of the natural numbers plus a unary relation U on N defined as follows.

For every $x \in \mathit{Var}$ and for some complete assignment h to $\mathcal{N}_U$,

(a) if $h(x) \in U$, then $\mathcal{N}_U \models U(x)[h]$;
(b) if $h(x) \notin U$, then $\mathcal{N}_U \models \neg U(x)[h]$.

We also stipulate the following convention.

(18) Convention: We assume that $\mathcal{L}_{basic}$ is ordered according to a fixed, recursive list l. Moreover, we assume that the first element of l is: $x_0 \dot{=} \overline{0}$.

Thus, in this section we refer to any sequence over $\mathcal{L}_{basic}$ as ordered according to l without further explanation. It will also be useful to use "world" and "structure" interchangeably.

We are now ready to start defining a basic agent $\langle \Psi^\circ, \mathbf{W}_\Psi\rangle$ from any given 01-agent Ψ, so as to be able to state and prove the following result. It gives the exact sense in which the paradigm of 01-coordination can be embedded into the m-coordination paradigm.

(19) Theorem: For all 01-agents Ψ,Φ, Ψ 01-coordinates with Φ iff $\langle \Psi^\circ, \mathbf{W}_\Psi\rangle$ m-coordinates with $\langle \Phi^\circ, \mathbf{W}_\Phi\rangle$.

[8] In a purely predicate framework without function and constant symbols, one might express the same ideas and results of this section by using an unary predicate symbol Z representing the singleton of 0 in place of $\overline{0}$, and a binary predicate symbol S representing the graph of the successor function in place of s.

In fact, our construction of $\langle \Psi^\circ, \mathbf{W}_\Psi \rangle$ from Ψ and of $\langle \Phi^\circ, \mathbf{W}_\Phi \rangle$ from Φ will satisfy a stronger property, namely:

(20) THEOREM: For all 01-agents Ψ,Φ, the following holds.

(a) If Ψ does not 01-coordinate with Φ, then for all $k \in N$, $_k|R(\Psi^\circ, \Phi^\circ)$ and $_k|R(\Phi^\circ, \Psi^\circ)$ are not environments for any world.

(b) If Ψ 01-coordinates with Φ, then for some $k \in N$, $_k|R(\Psi^\circ, \Phi^\circ)$ and $_k|R(\Phi^\circ, \Psi^\circ)$ are environments for a world in $\mathbf{W}_\Psi \cap \mathbf{W}_\Phi$.

Before proving Theorem (19), we give an informal account of Ψ°'s coordination behavior (Φ°'s behavior). By "behavior" here and in the sequel is meant a finite sequence of basic formulas produced in response to some interaction. For example, $R(\Psi^\circ, \Phi^\circ)|_n$ is the behavior of basic agent $\langle \Psi^\circ, \mathbf{W}_\Psi \rangle$ in response to basic agent $\langle \Phi^\circ, \mathbf{W}_\Phi \rangle$ for any $n \in N$. Here Ψ°'s behavior is required to make sense only on basic agents of the form $\langle \Phi^\circ, \mathbf{W}_\Phi \rangle$ with Φ being a 01-agent, because we make no claim in Theorem (19) about Ψ°'s behavior in response to other agents.

Let $U\langle \Psi, \Phi \rangle$ denote the set $\{n \in N \mid R(\Psi, \Phi)_n = 1\}$. Then:

- Ψ° gives some information to Φ° about Ψ's moves in response to Φ. This information is coded by literals of the form $U(x_n)$ or $\neg U(x_n)$, under the intuition that $U(x_n)$ codes the information "at stage n Ψ responds 1", and that $\neg U(x_n)$ codes the information "at stage n Ψ responds 0".

- Ψ° decodes the information coming from Φ° (again, coded by literals of the form $U(x_n)$ or $\neg U(x_n)$), about Φ's response to Ψ's moves.

- Ψ° describes an environment consisting of all literals true in $\mathcal{N}_{U\langle \Psi, \Phi \rangle}$ under the complete assignment h defined by $h(x_i) = i$ for all $i \in N$. At each stage, Ψ°'s description is partial, in the sense that only a finite part of $U\langle \Psi, \Phi \rangle$ is known to Ψ°. Even though at each step Ψ°'s knowledge of $U\langle \Psi, \Phi \rangle$ is partial, Ψ° updates his information according to the moves of Φ°.

- As soon as Ψ° discovers a disagreement between Ψ and Φ, that is, Ψ and Φ differ on some input, Ψ° outputs some (on the basis of the previous point) obviously false information, say $\neg x_0 \doteq \overline{0}$, and restarts the description of $\mathcal{N}_{U\langle \Psi, \Phi \rangle}$ by enumerating the basic diagram of $\mathcal{N}_{U\langle \Psi, \Phi \rangle}$ as before, using the assignment h to supply temporary names for the members of $\mathcal{N}_{U\langle \Psi, \Phi \rangle}$. The reason for this false information is that $\langle \Psi^\circ, \mathbf{W}_\Psi \rangle$ dislikes to coordinate with $\langle \Phi^\circ, \mathbf{W}_\Phi \rangle$ whenever Ψ does not 01-coordinate with Φ. If this happens, that is, if Ψ and Φ disagree infinitely often, or if Ψ and Φ's response sequences are finite, then Ψ° repeats its false information infinitely often. The result is that Ψ°'s description is a contradictory list of literals, with infinitely many elements of the form $x_0 \doteq \overline{0}$ and $\neg x_0 \doteq \overline{0}$. Thus, no possibility of coordination arises. Note that the false information $\neg x_0 \doteq \overline{0}$ is not misleading to Φ°, because he knows everything about $\mathcal{N}$ and the assignment h.

Recall that our goal is to define Ψ° and $\mathbf{W}_\Psi$ from bit agent Ψ (similarly: Φ° and $\mathbf{W}_\Phi$ from Φ) such that the following conditions hold.

(21) For all 01-agents Φ:

(*a*) If Ψ 01-coordinates with Φ, then for some $k \in N$, ${}_k|R(\Psi^\circ, \Phi^\circ)$ is an environment for $\mathcal{N}_{U\langle\Psi,\Phi\rangle} \in \mathbf{W}_\Psi$ and ${}_k|R(\Phi^\circ, \Psi^\circ)$ is an environment for $\mathcal{N}_{U\langle\Phi,\Psi\rangle} \in \mathbf{W}_\Phi$. Since the symmetric difference of $U\langle\Psi,\Phi\rangle$ and $U\langle\Phi,\Psi\rangle$ is finite, it follows that ${}_k|R(\Psi^\circ, \Phi^\circ)$ and ${}_k|R(\Phi^\circ, \Psi^\circ)$ are environments for some world in $\mathbf{W}_\Psi \cap \mathbf{W}_\Phi$.

(*b*) If Ψ does not 01-coordinate with Φ, then for all $k \in N$, both ${}_k|R(\Psi, \Phi)$ and ${}_k|R(\Phi, \Psi)$ contain infinitely many occurrences of a literal and its negation. Hence, for all $k \in N$, ${}_k|R(\Psi^\circ, \Phi^\circ)$ and ${}_k|R(\Phi^\circ, \Psi^\circ)$ are environments for no world.

We need some more notation. (Recall that for $n \in N$, $\sigma[n] = \langle \sigma_0 \cdots \sigma_n \rangle$.)

(22) DEFINITION: Let $\sigma \in BISEQ$ and 01-agent Ψ be given. We define $\Psi[\sigma] \in BISEQ$ as follows.

(*a*) $|\,\Psi[\sigma]\,| = |\,\sigma\,| + 1$;

(*b*) $\Psi[\sigma]_0 = \Psi(\emptyset)$.

(*c*) If $n < |\,\sigma\,|$ then $\Psi[\sigma]_{n+1} = \Psi(\sigma[n])$.

Definition (22) applies *mutatis mutandis* to basic agents and to sequences in *SEQ*.

(23) *Example:* Let $\sigma \in SEQ$ and $\Psi : SEQ \longrightarrow \mathcal{L}_{basic}$ be given. Suppose that $\sigma = \beta_0\beta_1\beta_2$. Then, $\Psi[\sigma] = \overline{\Psi(\sigma)}$ and $\Psi[\sigma] = \Psi[\sigma]_0\Psi[\sigma]_1\Psi[\sigma]_2\Psi[\sigma]_3$, where:

(*a*) $\Psi[\sigma]_0 = \Psi(\emptyset)$;

(*b*) $\Psi[\sigma]_1 = \Psi(\beta_0)$;

(*c*) $\Psi[\sigma]_2 = \Psi(\beta_0\beta_1)$;

(*d*) $\Psi[\sigma]_3 = \Psi(\beta_0\beta_1\beta_2)$.

The proof of Theorem (19) now proceeds through the construction of basic agents $\langle\Psi^\circ, \mathbf{W}_\Psi\rangle$, $\langle\Phi^\circ, \mathbf{W}_\Phi\rangle$ and a series of lemmas. First the construction.

Definition of $\mathbf{W}_\Psi$. Let 01-agent Ψ be given.

(24) DEFINITION: $\mathbf{W}_\Psi$ is the class of worlds $\mathcal{N}_U$ such that for some $\Phi \in \Lambda_0^1$,

(*a*) Ψ 01-coordinates with Φ, and

(*b*) $U \triangle U\langle\Psi,\Phi\rangle$ is finite.

For every $\sigma \in SEQ$, we now define the "environment transformation" τ, the environment e_σ, and communication function Ψ°.

Definition of Ψ°. $\tau(\sigma)$ is a compressed version of the information concerning the response sequence $R(\Psi, \Phi)$ coded by σ, and e_σ represents the partial description produced by Ψ° upon receiving σ and according to the recursive list *l* (see Convention (18)).

(25) DEFINITION: Let $\sigma \in SEQ$ be given. Denote by $S(\sigma)$ the set $\{n \in N \mid \text{for all}\, i \leq n\ (U(x_i) \in range(\sigma) \text{ or } \neg U(x_i) \in range(\sigma))\}$. The *e*-**transformation** is the total function $\tau : SEQ \longrightarrow BISEQ$ such that if $S(\sigma) = \emptyset$, then $\tau(\sigma) = \emptyset$. Otherwise, let m be the maximum of $S(\sigma)$. We define:

(*a*) $|\tau(\sigma)| = m + 1$.

(*b*) For $i < |\tau(\sigma)|$,

$$\tau(\sigma)_i = \begin{cases} 1 & \text{if there is}\, j < |\sigma|\ \sigma_j = U(x_i) \text{ and, for all}\, k \leq j,\ \sigma_k \neq \neg U(x_i) \\ 0 & \text{otherwise.} \end{cases}$$

Informally speaking, τ transforms the positive and negative information coded by a sequence in *SEQ* by means of a finite, binary sequence. For $\sigma \in SEQ$, observe that $\tau(\sigma)$ is generally shorter than σ.

(26) *Example:*

(*a*) Suppose that $\sigma = U(x_0)\beta_1\neg U(x_1)\beta_3\beta_4$. Then, $S(\sigma) = \{0, 1\}$ and $\tau(\sigma) = 10$. Let $\alpha_1 \in \mathcal{L}_{basic}$ neither of the form $U(x)$ nor $\neg U(x)$, and let $\alpha_2 = U(x_2)$. It follows that $S(\sigma\alpha_1) = \{0, 1\}$ and $\tau(\sigma\alpha_1) = 10$, thus $\tau(\sigma\alpha_1) = \tau(\sigma)$; while $S(\sigma\alpha_2) = \{0, 1, 2\}$ and $\tau(\sigma\alpha_2) = 101$, thus $\tau(\sigma\alpha_2) = \tau(\sigma)1$, *i.e.*, $\tau(\sigma\alpha_2) = \tau(\sigma)\tau(\sigma\alpha_2)_2$.

(*b*) Suppose that $\sigma = U(x_0)\beta_1\neg U(x_2)\beta_3\beta_4$. Then, $S(\sigma) = \{0\}$ and $\tau(\sigma) = 1$. Let $\alpha = \neg U(x_1)$. It follows that $S(\sigma\alpha) = \{0, 1, 2\}$ and $\tau(\sigma\alpha) = 101$, thus $\tau(\sigma\alpha) = \tau(\sigma)01$, *i.e.*, $\tau(\sigma\alpha) = \tau(\sigma)\tau(\sigma\alpha)_1\tau(\sigma\alpha)_2$.

(*c*) Suppose that $\sigma = \beta_0\neg U(x_2)\beta_2\beta_3\neg U(x_4)\beta_5 U(x_3)U(x_0)$. Then, $S(\sigma) = \{0, 1, 2, 3, 4\}$ and $\tau(\sigma) = 11010$. Let $\alpha = U(x_4)$. It follows that $S(\sigma\alpha) = S(\sigma)$ and $\tau(\sigma\alpha) = \tau(\sigma)$.

(*d*) Suppose that $\sigma = \beta_0 U(x_5)\beta_2 U(x_1)\neg U(x_4)\beta_5 U(x_3)\beta_7 U(x_0)$. Then, $S(\sigma) = \{0, 1\}$ and $\tau(\sigma) = 11$. Let $\alpha = U(x_2)$. Then $S(\sigma\alpha) = \{0, 1, 2, 3, 4, 5\}$ and $\tau(\sigma\alpha) = 111101$, so $\tau(\sigma\alpha) = \tau(\sigma)1101$, *i.e.* $\tau(\sigma\alpha) = \tau(\sigma)\tau(\sigma\alpha)_2 \ldots \tau(\sigma\alpha)_5$.

Note that case (*c*) of the example shows that if the conditional "if $U(x_i) \in range(\sigma)$" in Definition (25) were used in place of "if there is $j < |\sigma|\ (\sigma_j = U(x_i))$ and for all $k \leq j\ (\sigma_k \neq \neg U(x_i))$", then it would no longer be the case that $\tau(\sigma\alpha) = \tau(\sigma)$, because of $\tau(\sigma\alpha) = 11011$.

Definition (25) will be used in such a way that at any stage n in the coordination game between Ψ° and Φ°, when Ψ°'s last move is $R(\Psi^\circ, \Phi^\circ)_{n-1}$, σ is the finite sequence $R(\Psi^\circ, \Phi^\circ)|_{n-1}$ of Φ°'s previous responses to Ψ°. We will show pretty soon that in this very special case $|\tau(\sigma^-)| = |\tau(\sigma)| - 1$. The reader should keep this in mind in reading Lemma (27) and next two definitions.

Example (26) sets some general properties of transformation τ. We state these properties as follows.

(27) LEMMA: Let $\sigma \in SEQ$ be given.

(a) For all $\eta \in SEQ$, if $\sigma \sqsubseteq \eta$ then $\tau(\sigma) \sqsubseteq \tau(\eta)$.

(b) Let $|\sigma| = n+1$ with $n \in N$. Then, $\tau(\sigma)$
$= \tau(\sigma^-)$, if for all $i \in N$, $\sigma_n \neq U(x_i)$ and $\sigma_n \neq \neg U(x_i)$, or either $\sigma_n = U(x_i)$ or $\sigma_n = \neg U(x_i)$ for some $i \in N$ with $i < |\tau(\sigma^-)|$;
$= \tau(\sigma^-)\tau(\sigma)_{|\tau(\sigma^-)|} \ldots last(\tau(\sigma))$, otherwise.

(28) DEFINITION: Let $\sigma \in SEQ$, $\Psi \in \Lambda^1_0$, and complete assignment h to $\mathcal{N}$ be given. Suppose that $h(x_i) = i$ for all $i \in N$. The **p-environment for $\mathcal{N}$ and h generated by Ψ on** $\tau(\sigma)$ is the sequence e_σ such that, if $\Psi[\tau(\sigma)]$ is undefined, then e_σ is undefined. Otherwise, $range(e_\sigma) = \{\beta \in \mathcal{L}_{basic} \,|\, \mathcal{N} \models \beta[h]\} \cup \{U(x_i) \,|\, i < |\tau(\sigma)| \text{ and } \Psi[\tau(\sigma)]_i = 1\} \cup \{\neg U(x_i) \,|\, i < |\tau(\sigma)| \text{ and } \Psi[\tau(\sigma)]_i = 0\}$.

In particular, we note that e_σ strictly depends on list l (see Convention (18)) and on assignment h. Moreover, it follows from Convention (18) that $x_0 \doteq \overline{0}$ is the first element of e_σ. This is indeed crucial in the next definition.

(29) DEFINITION: Let $\sigma \in SEQ$ and $\Psi \in \Lambda^1_0$ be given. Ψ° is the communication function such that if $\Psi[\tau(\sigma)]$ is undefined, then $\Psi^\circ(\sigma)$ is undefined. Otherwise, let i_0 be the maximum $i < |\Psi^\circ[\sigma^-]|$ such that $\Psi^\circ[\sigma^-]_i$ is of the form $\neg x_0 \doteq \overline{0}$; if such an i does not exist, let $i_0 = -1$. Define:

(a) if $\Psi(\tau(\sigma^-)) = (\neg x_0 \doteq \overline{0})$, then $\Psi^\circ(\sigma)$ is of the form $x_0 \doteq \overline{0}$;
(b) if either $\tau(\sigma) = \tau(\sigma^-)$ or $\tau(\sigma) \neq \tau(\sigma^-)$ and $last(\tau(\sigma)) = \Psi(\tau(\sigma^-))$, $\Psi^\circ(\sigma)$ is the first element β in e_σ such that for all $i_0 < i < |\Psi^\circ[\sigma^-]|$, $\beta \neq \Psi^\circ[\sigma^-]_i$;
(c) if $\tau(\sigma) \neq \tau(\sigma^-)$ and $last(\tau(\sigma)) \neq \Psi(\tau(\sigma^-))$, $\Psi^\circ(\sigma)$ is of the form $\neg x_0 \doteq \overline{0}$.

We say that Ψ° is the **first-order equivalent of** Ψ.

(30) *Remark:* Note that $\Psi^\circ[\sigma^-]$ in Definition (29) is the sequence of the previous moves of Ψ°, and i_0 represents the last place where an error occurred. We can imagine that Ψ°'s description of his actual world restarts from the place $i = i_0 + 1$. Thus, if $i_0 = -1$, the previous definition suggests that the agents' coordination process starts again from the beginning (place $i = 0$). In Definition (29), clause (a) says that if the last literal of σ does not contain any new information about Φ's moves, or if the last literal of σ contains information about a new move of Φ, but this move agrees with Ψ's last move, then Ψ° outputs on σ a new literal in e_σ, namely, the first element of e_σ that is not output after the last error. According to Definition (29)(b), Ψ° response on σ is a failure, *i.e.* a literal of the form $\neg x_0 \doteq \overline{0}$, if the last literal in σ contains information about a new move of Φ and this move differs from Ψ's last move. Finally, clause (c) ensures that $x_0 \doteq \overline{0}$ is to be played by Ψ° whenever 01-agents Ψ and Φ disagree from the beginning.

Definition (29) concludes the construction of Ψ°. Observe:

(31) LEMMA: For all $\Psi \in \Lambda_0^1$:

(a) If Ψ is total, then Ψ° is total.
(b) If Ψ is recursive, then Ψ° is recursive.

The next lemma is a reformulation of Lemma (27) for sequences of basic formulas "played" by agents Ψ° and Φ°. Before stating the lemma, we note that from the previous three definitions it follows that for every sequence σ of the form $R(\Psi^\circ, \Phi^\circ)|_n$, there is $j < |\sigma|$ such that both $\sigma_j = U(x_i)$ and $\sigma_k \neq \neg U(x_i)$ for all $k \leq j$ if and only if $U(x_i) \in \mathit{range}(\sigma)$. This is will be indeed useful in the proof of the lemma, where we focus for convenience on the shortest notation.

(32) LEMMA: For all $\Psi, \Phi \in \Lambda_0^1$ such that $R(\Psi, \Phi)$ is an infinite sequence, the following holds. For all $n > 0$, either

(a) $\tau(R(\Psi^\circ, \Phi^\circ)|_n) = \tau(R(\Psi^\circ, \Phi^\circ)|_{n-1})$, or
(b) $\tau(R(\Psi^\circ, \Phi^\circ)|_n) = \tau(R(\Psi^\circ, \Phi^\circ)|_{n-1})\mathit{last}(\tau(R(\Psi^\circ, \Phi^\circ)|_n))$.

Proof: We have only to show that if $\tau(R(\Psi^\circ, \Phi^\circ)|_n) \neq \tau(R(\Psi^\circ, \Phi^\circ)|_{n-1})$, then $|\tau(R(\Psi^\circ, \Phi^\circ)|_{n-1})| = |\tau(R(\Psi^\circ, \Phi^\circ)|_n)| - 1$. (The lemma then follows immediately as a special case of Lemma (27)(b).) Suppose $\tau(R(\Psi^\circ, \Phi^\circ)|_n) \neq \tau(R(\Psi^\circ, \Phi^\circ)|_{n-1})$. Then by the definition of τ, either $R(\Psi^\circ, \Phi^\circ)_{n-1} = U(x_i)$ or $R(\Psi^\circ, \Phi^\circ)_{n-1} = \neg U(x_i)$ with $i = |\tau(R(\Psi^\circ, \Phi^\circ)|_{n-1})|$. From the definition of Ψ° it follows that either $\tau(R(\Psi^\circ, \Phi^\circ)|_n) = \tau(R(\Psi^\circ, \Phi^\circ)|_{n-1})1$ or $\tau(R(\Psi^\circ, \Phi^\circ)|_n) = \tau(R(\Psi^\circ, \Phi^\circ)|_{n-1})0$. Hence, $|\tau(R(\Psi^\circ, \Phi^\circ)|_n)| = |\tau(R(\Psi^\circ, \Phi^\circ)|_{n-1})| + 1$. ■

(33) LEMMA: For all $\Psi, \Phi \in \Lambda_0^1$ such that $R(\Psi, \Phi)$ is an infinite sequence, the following conditions hold.

(a) For all $n \in N$, $\tau(R(\Psi^\circ, \Phi^\circ)|_n) \sqsubseteq R(\Psi, \Phi)|_n$.
(b) For all $n \in N$, there is $m \in N$ such that $R(\Psi, \Phi)|_n \sqsubseteq \tau(R(\Psi^\circ, \Phi^\circ)|_m)$.

Thus, Lemma (33)(a) says that Φ° receives only correct information from Ψ° about Ψ's moves, or also that the (positive and negative) information contained in $R(\Psi^\circ, \Phi^\circ)$ reflects the (positive and negative) information contained in $R(\Psi, \Phi)$. Lemma (33)(b) says that sooner or later Φ° receives from Ψ° all information about $R(\Psi, \Phi)$. In other words, the (positive and negative) information contained in $R(\Psi, \Phi)$ will appear in $R(\Psi^\circ, \Phi^\circ)$ in finite time.

Proof of Lemma (33): (33)(a) By induction on n. For $n = 0$ the claim holds because of $\tau(\emptyset) = \emptyset$. For $n > 0$, assume $\tau(R(\Psi^\circ, \Phi^\circ)|_{n-1}) \sqsubseteq R(\Psi, \Phi)|_{n-1}$ and $\tau(R(\Phi^\circ, \Psi^\circ)|_{n-1}) = R(\Phi, \Psi)|_{n-1}$ (inductive hypothesis). Let $\sigma = R(\Psi^\circ, \Phi^\circ)|_n$ and $i = |\tau(\sigma)|$. We want to prove that $\tau(\sigma) \sqsubseteq R(\Psi, \Phi)|_n$. By Lemma (32), either $\tau(\sigma) = \tau(\sigma^-)$ or $\tau(\sigma) = \tau(\sigma^-)\tau(\sigma)_{i-1}$. We consider the following three cases:

Case 1: $\tau(\sigma) = \tau(\sigma^-)$. It follows by the inductive hypothesis that $\tau(\sigma) \sqsubseteq R(\Psi, \Phi)|_{n-1}$. Observe that $R(\Psi, \Phi)|_n = R(\Psi, \Phi)|_{n-1} R(\Psi, \Phi)_{n-1}$, and then $R(\Psi, \Phi)|_{n-1} \sqsubseteq R(\Psi, \Phi)|_n$. Hence $\tau(\sigma) \sqsubseteq R(\Psi, \Phi)|_n$.

Case 2: $\tau(\sigma) \neq \tau(\sigma^-)$ and $\tau(\sigma)_{i-1} = 1$. Then, by the definition of τ, $U(x_{i-1}) \in range(\sigma)$, that is $U(x_{i-1})$ is one of the $R(\Psi^\circ, \Phi^\circ)_k$'s, $k = 0, 1, ..., n-1$. Because of $R(\Psi^\circ, \Phi^\circ)_{n-1} = \Psi^\circ(R(\Phi^\circ, \Psi^\circ)|_{n-1})$, by the definition of Ψ° it follows that $U(x_{i-1}) \in range(e_{R(\Phi^\circ, \Psi^\circ)|_{n-1}})$ and, by the definition of $e_{R(\Phi^\circ, \Psi^\circ)|_{n-1}}$, that $\Psi[\tau(R(\Phi^\circ, \Psi^\circ)|_{n-1})]_{i-1} = 1$. By the inductive hypothesis and by the definition of $\Psi[\cdot]_{i-1}$, it follows that $\Psi(R(\Phi, \Psi)|_{n-1}) = 1$ (observe: $n = i - 1$). Hence, $R(\Psi, \Phi)_{n-1} = 1$. We have thus obtained: $\tau(\sigma) = \tau(\sigma^-) R(\Psi, \Phi)_{n-1}$. It follows by the inductive hypothesis that $\tau(\sigma) \sqsubseteq R(\Psi, \Phi)|_n$.

Case 3: $\tau(\sigma) \neq \tau(\sigma^-)$ and $\tau(\sigma)_{i-1} = 0$. By the definition of τ, $\tau(\sigma)_{i-1} = 0$ implies $U(x_{i-1}) \notin range(\sigma)$, that is $U(x_{i-1})$ is none of the $R(\Psi^\circ, \Phi^\circ)_k$'s, $k \leq n-1$. Then, by the definition of Ψ°, $R(\Psi^\circ, \Phi^\circ)_{n-1} = \neg U(x_{i-1})$ and $\neg U(x_{i-1}) \in range(e_{R(\Phi^\circ, \Psi^\circ)|_{n-1}})$, thus $\Psi[\tau(R(\Phi^\circ, \Psi^\circ)|_{n-1})]_{i-1} = 0$. By the inductive hypothesis and by the definition of $\Psi[\cdot]_{i-1}$, it follows that $\Psi(R(\Phi, \Psi)|_{n-1}) = 0$. Hence, $R(\Psi, \Phi)_{n-1} = 0$. We have thus established: $\tau(\sigma) = \tau(\sigma^-) R(\Psi, \Phi)_{n-1}$. By the inductive hypothesis, it follows that $\tau(\sigma) \sqsubseteq R(\Psi, \Phi)|_n$. This concludes the proof of (33)(*a*).

(33)(*b*) By induction on n. For $n = 0$ the claim holds for every $m \in N$. For $n > 0$, assume that for some $m \in N$, $R(\Psi, \Phi)|_{n-1} \sqsubseteq \tau(R(\Psi^\circ, \Phi^\circ)|_m)$ and $R(\Phi, \Psi)|_{n-1} = \tau(R(\Phi^\circ, \Psi^\circ)|_{m+1})$ (inductive hypothesis). For short, let $\sigma = R(\Psi^\circ, \Phi^\circ)|_{m+1}$ and $i = |\tau(\sigma)|$. We prove that $R(\Psi, \Phi)|_n \sqsubseteq \tau(\sigma)$.

Observe that $R(\Psi, \Phi)|_n = R(\Psi, \Phi)|_n R(\Psi, \Phi)_{n-1}$. We consider two cases:

Case 1: $R(\Psi, \Phi)_{n-1} = 1$. By the definition of $R(\Psi, \Phi)$, $\Psi(R(\Phi, \Psi)|_{n-1}) = 1$. By the inductive hypothesis, it follows that $\Psi(\tau(R(\Phi^\circ, \Psi^\circ)|_{m+1})) = 1$. Observe that $|\tau(R(\Phi^\circ, \Psi^\circ)|_{m+1})| = n-1$, thus $\Psi[\tau(R(\Phi^\circ, \Psi^\circ)|_{m+1})]_n = 1$. This means that $U(x_n) \in range(e_{R(\Phi^\circ, \Psi^\circ)|_{m+1}})$; by the definition of Ψ° it follows that $U(x_n) = \Psi^\circ(R(\Phi^\circ, \Psi^\circ)|_{m+1})$, namely, $U(x_n) = R(\Psi^\circ, \Phi^\circ)_m$. Then $U(x_n) \in range(\sigma)$ and, by the definition of τ, $\tau(\sigma)_n = 1$. We have thus obtained: $R(\Psi, \Phi)_{n-1} = \tau(\sigma)_n$. Since $R(\Psi, \Phi)|_n = R(\Psi, \Phi)|_n R(\Psi, \Phi)_{n-1}$, it follows that $R(\Psi, \Phi)|_n \sqsubseteq \tau(\sigma^-)\tau(\sigma)_n$, by the inductive hypothesis. Then, by Lemma (32), $R(\Psi, \Phi)|_n \sqsubseteq \tau(\sigma)$.

Case 2: $R(\Psi, \Phi)_{n-1} = 0$. Similarly to the proof of case 1 we obtain that $\Psi[\tau(R(\Phi^\circ, \Psi^\circ)|_{m+1})]_n \neq 1$, with $|\tau(R(\Phi^\circ, \Psi^\circ)|_{m+1})| = n-1$. This means that $U(x_n) \notin range(e_{R(\Phi^\circ, \Psi^\circ)|_{m+1}})$. By the definition of Ψ° it follows that $U(x_n) \neq \Psi^\circ(R(\Phi^\circ, \Psi^\circ)|_{m+1})$, hence $U(x_n) \neq R(\Psi^\circ, \Phi^\circ)_m$. Then $U(x_n) \notin range(\sigma)$ (because of $n - 1 \leq m$), and, by the definition of τ, $\tau(\sigma)_n = 0$. We have thus obtained: $R(\Psi, \Phi)_{n-1} = \tau(\sigma)_n$. Since $R(\Psi, \Phi)|_n = R(\Psi, \Phi)|_n R(\Psi, \Phi)_{n-1}$, by the inductive hypothesis it follows that $R(\Psi, \Phi)|_n \sqsubseteq \tau(\sigma^-)\tau(\sigma)_n$. Then, by Lemma (32), $R(\Psi, \Phi)|_n \sqsubseteq \tau(\sigma)$. ■

(34) COROLLARY: For all $\Psi, \Phi \in \Lambda^1_0$, $R(\Psi, \Phi) = R(\Phi, \Psi)$ iff $R(\Psi^\circ, \Phi^\circ) = R(\Phi^\circ, \Psi^\circ)$.

The next lemma says that, if $R(\Psi, \Phi)$ is not an infinite sequence, neither is $R(\Psi^\circ, \Phi^\circ)$.

(35) LEMMA: For all $\Psi, \Phi \in \Lambda_0^1$, if for some $n \in N$ $R(\Psi, \Phi)_n$ is undefined, then for some $m \in N$ $R(\Psi^\circ, \Phi^\circ)_m$ is undefined.

Proof: Suppose that $R(\Psi, \Phi)_n$ undefined for some $n \in N$ and let $R(\Phi, \Psi)|_n = \tau(\sigma)$ for some $\sigma \in SEQ$. By Lemma (33), $\sigma = R(\Phi^\circ, \Psi^\circ)|_m$ for some $m \in N$. Then, $R(\Psi, \Phi)_n = \Psi(\tau(R(\Phi^\circ, \Psi^\circ)|_m))$. By the definition of Ψ°, it follows that $\Psi^\circ(R(\Phi^\circ, \Psi^\circ)|_m) = R(\Psi^\circ, \Phi^\circ)_m$ is undefined. ■

Proof of Theorem (19): Let $\Psi, \Phi \in \Lambda_0^1$ be given. For the left-to-right direction, suppose that Ψ 01-coordinates with Φ. Then $R(\Psi^\circ, \Phi^\circ)$ and $R(\Phi^\circ, \Psi^\circ)$ contain only finitely many errors. (An error occurs in correspondence to a disagreement between Ψ and Φ, and only finitely many disagreements occur.) After the last error, Ψ° and Φ° start describing their worlds: Ψ°'s describes $\mathcal{N}_{U\langle\Psi,\Phi\rangle}$ by enumerating via l a subset of all literals true in $\mathcal{N}_{U\langle\Psi,\Phi\rangle}$ under the assignment h. By Lemma (33), only true literals become elements of Ψ°'s description, and all true literals enter in it. Since the order of literals is always the one induced by the list l, if a literal β is in Ψ°'s description, then sooner or later it is output by Ψ°. Indeed, as soon as β enters in Ψ°'s description, by construction only literals preceding it in the list l can be output by Ψ°. As soon as all true literals preceding β are output, at the next stage β is output by Ψ°. Thus, there exists $k \in N$, namely, the number of literals occurring in $R(\Psi^\circ, \Phi^\circ)$ before the last error, last error included, such that $_k|R(\Psi^\circ, \Phi^\circ)$ is an environment for $\mathcal{N}_{U\langle\Psi,\Phi\rangle}$ and $_k|R(\Phi^\circ, \Psi^\circ)$ is an environment for $\mathcal{N}_{U\langle\Phi,\Psi\rangle}$. Since Ψ 01-coordinates with Φ, the symmetric difference between $U\langle\Psi, \Phi\rangle$ and $U\langle\Phi, \Psi\rangle$ is finite. It follows that both $\mathcal{N}_{U\langle\Psi,\Phi\rangle}$ and $\mathcal{N}_{U\langle\Phi,\Psi\rangle}$ belong to $\mathbf{W}_\Psi \cap \mathbf{W}_\Phi$. Hence, $\langle\Psi^\circ, \mathbf{W}_\Psi\rangle$ and $\langle\Phi^\circ, \mathbf{W}_\Phi\rangle$ m-coordinate.
For the right-to-left direction, we claim that if Ψ does not 01-coordinate with Φ then $\langle\Psi^\circ, \mathbf{W}_\Psi\rangle$ and $\langle\Phi^\circ, \mathbf{W}_\Phi\rangle$ do not m-coordinate. To prove the claim, suppose that Ψ does not 01-coordinate with Φ. Two cases arise:

Case 1: Either $R(\Psi, \Phi)$ or $R(\Phi, \Psi)$ is not an infinite sequence. Then the claim follows directly from Lemma (35).

Case 2: Both $R(\Psi, \Phi)$ and $R(\Phi, \Psi)$ are infinite sequences and $R(\Psi, \Phi)_k \neq R(\Phi, \Psi)_k$ for infinitely many $k \in N$. Observe that, by Convention (18), $x_0 \doteq \overline{0}$ is the first element of the list l. Then, the construction of $\langle\Psi^\circ, \mathbf{W}_\Psi\rangle$ and $\langle\Phi^\circ, \mathbf{W}_\Phi\rangle$ ensures that in $R(\Psi^\circ, \Phi^\circ)$ and $R(\Phi^\circ, \Psi^\circ)$ there are infinitely many occurrences of $x_0 \doteq \overline{0}$ and $\neg x_0 \doteq \overline{0}$. Thus, neither of $R(\Psi^\circ, \Phi^\circ)$ and $R(\Phi^\circ, \Psi^\circ)$ is an environment for any world, not even after deleting finitely many literals. So, $\langle\Psi^\circ, \mathbf{W}_\Psi\rangle$ and $\langle\Phi^\circ, \mathbf{W}_\Phi\rangle$ do not m-coordinate. ■

6 Conclusion

We briefly presented a recursion theoretic paradigm of binary coordination equivalent to the paradigm of "learning to coordinate choices" of [MO99], where two players repeatedly interact in order to stabilize in the limit over two possible behaviors. We then advanced and discussed a model-theoretic paradigm of pairwise coordination. In our paradigm the agents preferences and beliefs are expressible within a first-order framework. As the main result of this paper, binary coordination was shown to be reducible to a special case of model-theoretic coordination. As a consequence, Montagna and Osherson's paradigm is proved to be a special case of our paradigm of model-coordination, in the sense that two 01-agents coordinate if and only if their first-order equivalent basic agents coordinate in our model-theoretic setting. An important difference between model-coordination and binary coordination is that in the former the agents preferences and beliefs can be modelled.

7 Acknowledgements

A preliminary version of this article is [AdJM00]. The first author thanks the ILLC staff for kindly support during his research visit in 1999.

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.

[Ago01] A. Agostini. *Paradigms of Coordination and Solvability.* PhD thesis, LOMIT-XI, Dip. di Matematica, Università di Siena, March 2001.

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

[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.

[Kei77] H. J. Keisler. Fundamentals of model theory. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 47–104. North-Holland Publishing Company, 1977.

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

[MO99] F. Montagna and D. Osherson. Learning to coordinate: A recursion theoretic perspective. *Synthese*, 118(3):363–382, 1999.

[OR94] M. J. Osborne and A. Rubinstein. *A Course in Game Theory.* The MIT Press, Cambridge, MA, 1994.

[OW86] D. Osherson and S. Weinstein. Identification in the limit of first order structures. *Journal of Philosophical Logic*, 15:55–81, 1986.