

# Notes on formalizing coordination\*

Alessandro Agostini

Università di Siena  
LOMIT - Dipartimento di Matematica  
via del Capitano, 15  
53100 SIENA  
Italy  
agostini@unisi.it

**Abstract.** This paper concerns with 2-agents coordination games—we call them *paradigms of coordination*. To coordinate, agents' behaviour must eventually stabilize to a set of basic formulas that express a suitable part of agents' "nature". Four paradigms are advanced and discussed. Several new perspectives are provided to coordinating agents. Coordination via belief revision and cooperation by team work are two.

## 1 Introduction

A real problem in extending single agent systems to cooperative settings is determining methods of *coordination*. However, agents may be based on different languages, they may have different behaviours, they may use different representations of information and reasoning strategies, and they may have different interaction capabilities. We will refer to all this as to *the coordination problem*. Investigations into the coordination problem can be divided into three general classes: those based on convention, those based on communication, and those based on learning. Some example in the first class was given in AI by Shoham and Tennenholtz [25, 24], where some social laws or 'conventions' [11] are imposed by the system designer (see also [20]) so that optimal joint action is assured. In the second class, agents' coordination is based on communication (see for instance [28]). This second class might be thought as a special case of the normative class, where the communication language is assumed to be the convention. So, what makes this class different is rather the emphasis it gives to the communicative agents' skills with regard to agents' coordination problems. In this class, it does make sense to speak about *failure messages* that prevent the agents from coordinating (see for instance [30] for some further remarks and references on the influence of the *speech act theory* in communication). In the third class, coordination might be learned through repeated interaction (see for instance [2, 22,

---

\* I thank Franco Montagna and Dick de Jongh for fundamental feedback on the first draft of this article and for research guidance. I am indebted to Daniel Osherson for early fruitful discussions on coordination and to Fausto Giunchiglia for invaluable suggestions. This work has been done at ILLC, Amsterdam. I kindly thank ILLC for the excellent research environment.

29] and the bibliography cited). It is evident that the coordination problem is central in several contemporary disciplines like Sociology, Artificial Intelligence, Game Theory, Computer Science. We shall not here attempt to synthesize this vast literature save to remark that a general understanding of the conditions under which coordination can be achieved, and exactly how it relates to agents' learning ability and their background knowledge and beliefs, are problems that have not yet been thoroughly explored.

Our ultimate goal is to provide a framework and a methodology which will allow us to investigate the coordination problem from a *learning-by-discovery* [29] perspective. To achieve this goal, we need a suitable formalism. We focus on the model-theoretic tradition of *Formal Learning Theory*—say [23, 7, 19, 18, 12], that descends from the pioneering studies on inductive inference developed by [27, 21, 8, 1]. The work in the recursive-theoretic tradition concerns algorithms for inferring recursive functions from finite samples of their graphs, and has been adapted successively to characterize abstract languages in the limit. The model-theoretic tradition is more recent; its main aim is to provide a formal framework for learning first order theories and models. The recursive-functional approach to learning has been extended to characterize recursive functions by means of coordination in the limit. A natural question is whether a similar framework can be developed to investigate the coordination problem. What follows is an attempt to present such a conception, and to describe what place coordination has within it. There are many definitions, some remarks, no theorems, and a great deal left to be investigated.

Our discussion begin in Section 2 with an intuitive idea of the coordination problem we are thinking on. In Section 3 we then proceed by presenting the logical framework. Four paradigms of coordination are advanced. We conclude in Section 4 with a discussion of several issues that promise to make the boorderline between formal learning theory and agent theory an exciting area of research for the foreseeable future.

## 2 Coordination as a 2-agents game

We start with a concrete example intended to help the reader interpret some of the abstract concepts described later. Thus, we can image two “worlds-based agents”, say **Alfonso** and **Barbara**, whose “background world” is represented by two nonempty classes of structures **A** and **B** of a given signature, and whose aim is to coordinate, *e.g.* to solve a common problem. For doing this, **Alfonso** and **Barbara** try to communicate with each other in order to respectively end up with a description of two structures  $\mathcal{A} \in \mathbf{A}$  and  $\mathcal{B} \in \mathbf{B}$  such that  $\mathcal{A}$  is sufficiently close to **B** and  $\mathcal{B}$  is sufficiently close to **A**. We expect that the more **Alfonso** is like **Barbara**, the better chance **Alfonso** and **Barbara** have of reaching coordination. To dramatize, let us suppose that each agent does not know the “background world” of the other agent, and that agents were never before in a similar situation, so they cannot rely on past experience to solve their coordination problem—no common knowledge arises between agents despite of their

common language. This “drama” is indeed a basic ingredient of what we called the coordination problem. We also image that the agents are “rational” (*e.g.* in their “communication ability”) and both know that joint cooperation is better than joint defection, but each has no idea what sort of agents the opponent is. **Alfonso** and **Barbara**’s decisions have a strategic component. Since strategic interactions are best modeled by game theory, we image the agents’ coordination problem as a form of 2-players game, which we qualify as a *coordination game*.

To start the game, **Alfonso** is conceived as choosing one member from **A** to be his “actual world”; **Alfonso**’s choice is initially unknown to **Barbara**. **Alfonso** then provides a “clue” (the **Alfonso**’s *behaviour*) about his world. **Barbara** does her choice as well (we image the game is *synchronous*, *i.e.* both agents are not taking turns but rather are choosing simultaneously), and provides **Alfonso** with a clue about her actual world. We can assume that **Alfonso** and **Barbara** are allowed to change their actual world at each step of the game, provided that they remain coherent with the behaviour they have shown since then. Clearly, it is safe to begin with a behaviour coherent on many worlds in the class. In this way, if say **Alfonso** realizes that the structure he has in mind, *i.e.* his actual world, is not close enough to **Barbara**’s world, he can change it, and vice versa. Agents may provide “bad clues”, in principle. Thus, we can consider a paradigm of *coordination-by-failure* according to which an agent can give to the other agent a ‘failure message’. In such a case, the agents are allowed to start the game again from the beginning. Of course, to reach coordination this should occur only finitely often. **Alfonso**’s clues constitute the data upon which **Barbara** will base her hypotheses on **Alfonso**’s background world, that eventually become themselves a clue for **Alfonso** about **Barbara**’s world. And so forth. Each time **Alfonso** provides a new clue, **Barbara** may produce a new hypothesis, and a new clue for **Alfonso** as well. **Alfonso** and **Barbara** *win the game*—and we say that they solve their coordination problem, if the successive clues about their own background worlds eventually stabilize to a consistent set of hypotheses satisfiable in one of the other agent’s worlds. Both lose otherwise. As a necessary condition for winning, each agent’s behaviour must be consistent with agent’s own background world.

### 3 Logical framework

Five concepts figure in the foregoing game-theoretic picture of coordination: worlds, agents, clues, descriptions, success. We formalize them as follows.

*Notation.* We fix a (countable, decidable) first-order language  $\mathcal{L}_{form}$  with vocabulary  $\mathcal{L}$  and countable set of variables  $Var$ . Unless stated otherwise,  $\mathcal{L}$  and  $Var$  will remain fixed. We use  $\mathcal{L}_{sen}$  and  $\mathcal{L}_{basic}$  to denote, respectively, the set of sentences (or *closed formulas*, that is no free variables occur) and the set of literals (or *basic formulas*) 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  $SEQ$ . Some further notation is as follows. The set  $\{0, 1, 2, \dots\}$  of natural numbers is denoted by  $N$ . If  $X$  is a set,  $pow(X)$  is the set of all subsets of  $X$  and  $X^\omega$  is the set of

infinite sequences over  $X$ . A sequence in  $X^\omega$  is called an  $\omega$ -sequence (over  $X$ ). Let  $\tau$  be an  $\omega$ -sequence. We write  $\tau(i)$ ,  $i \in \mathbb{N}$ , for the finite sequence  $\langle \tau_0 \dots \tau_i \rangle$ , and  $\tau|_i$  for the proper initial segment of length  $i$  in  $\tau$ . We write  $length(\eta)$  for the length of a finite sequence and  $\eta_i$  for the  $i$ th element of  $\eta$ ,  $0 \leq i < length(\eta)$ . We write  $range(\eta)$  for the set of elements of any sequence. We denote the finite sequence of length zero by  $\emptyset$ .

Otherwise, our semantic notions are standard.<sup>1</sup> In particular,  $\mathcal{L}$ -structure  $\mathcal{S}$  is a model of  $\Gamma \subseteq \mathcal{L}_{form}$ , and  $\Gamma$  is said to be *satisfiable in*  $\mathcal{S}$ , if there is an assignment  $h : \text{Var} \rightarrow |\mathcal{S}|$  with  $\mathcal{S} \models \Gamma[h]$ .  $\Gamma$  is *satisfiable* if it is satisfiable in some structure. The class of models of  $\Gamma$  is denoted:  $MOD(\Gamma)$ .

### 3.1 Worlds

We begin to give substance to our view of coordination by representing the possible realities, or “worlds”, where the coordination problem may arise. By *world* we shall here mean any countable structure that interprets  $\mathcal{L}$ . Worlds may be conceived as the “possible truths” for the agents. We are interested in aggregations of such worlds, namely, countable collections of worlds. These collections may be intuitively thought as the set of realities of a given agent. To see how, we must first say what we mean by an “agent”.

### 3.2 Agents

What has to be termed *agent* and what does not is a long debate in Artificial Intelligence (see for instance [5]; also [30] for a survey). Here we do not try a full explanation of our conception of “agent” from this more general perspective. In the sequel, rather, agents are conceived as systems that examine (partial) evidence coming from other agents’ behaviours or empirical data and emit hypotheses and clues. Agents are possibly bounded-resource systems and can fail on some input. We shall address some question about bounded-resource agents later in this paper, with the proviso that a deeper discussion of “real”, say computable, agents is to come.

To formalize the mixture of data and failures we need, let us extend  $\mathcal{L}$  with the symbol  $\perp$  for ‘message failure’. We will thus be interested in the collection of all the *finite* sequences whose elements are basic formulas and  $\perp$ . We denote such collection: *FSEQ*. Any member  $\sigma$  of *FSEQ* is to be interpreted as a finite evidence available to agents at time  $t = length(\sigma)$ . Thus,  $\sigma$  may be thought as an “evidential status” or a “situation” that recapitulate the information available to agents about an underlying world at a certain moment of observation. Note that *FSEQ* is countable, because of  $\mathcal{L}_{basic}$  does. We now record the official definition of “agent”.

**Definition 1.** *A (basic) agent is any mapping from FSEQ to  $\mathcal{L}_{basic} \cup \{\perp\}$ .*

---

<sup>1</sup> See for instance [14] for a standard reference.

We say that  $\mathcal{L}$  is the *agent’s language*. A basic agent might be partial or total, recursive or nonrecursive. Thus, agents examine data recorded in a finite representation and emit hypotheses or “clues” about the world to be represented by the data, or also they “suspend the judgement” by saying  $\perp$ . For  $\sigma \in FSEQ$  being the input of agent  $\Psi$ ,  $\Psi(\sigma)$  represents  $\Psi$ ’s behaviour with respect to the sequence  $\sigma$  of facts observed. In particular,  $\Psi(\sigma) = \perp$  may be interpreted as  $\Psi$ ’s “suspension of coordination” if  $\sigma$  collects the clues from some agent’s output, and as “suspension of the judgement” if  $\sigma$  collects data elsewhere, say from Nature.

Agents as functions are not enough. Indeed, agents of Definition 1 do not capture the basic ingredients of agency (see for instance, [30] and the references cited there). Moreover, according to the intuitive picture drawn in Section 2, when faced with any coordination problem an agent is conceived as trying to coordinate to the other agent advancing successive clues about his or her own “background world”. If an agent realizes that his “actual world” (*i.e.* the world he “has in mind”) is not close enough to some world of the other agent, he can change his actual world or break off the coordination process by playing  $\perp$ . To state all this precisely, basic agents must be restricted to “worlds-based agents”. We rely on the following definition.

**Definition 2.** *Let  $\Psi_0$  be a basic agent and  $\mathbf{A}$  be a nonempty class of worlds. We say that  $\Psi = \langle \Psi_0, \mathbf{A} \rangle$  is a worlds-based agent.*

For all  $\sigma \in SEQ$ , we then write  $\Psi(\sigma)$  for  $\Psi_0(\sigma)$ . Moreover, to shorten the terminology we allow ourselves to say “ $\langle \Psi_0, \mathbf{A} \rangle$  is an agent” in place of “ $\langle \Psi_0, \mathbf{A} \rangle$  is a worlds-based agent”. The class of all such worlds-based agents is denoted:  $\Delta$ . As basic agents, worlds-based agents may be computable or noncomputable. Of the two components of  $\Psi$ ,  $\Psi_0$  is said to be the *communication ability of  $\Psi$*  and  $\mathbf{A}$  is said to be the *background world of  $\Psi$* . We also say that  $\Psi$  is *based on  $\mathbf{A}$* . To fix intuitions one might think of a background world as representing the agent’s belief space. In this case, background worlds generalize the scientist’s “mono-world” habitat of the first-order paradigm of inquiry [13]. Thus, the scientist’s question: “What is true in my world?” (cf. [13], p. 63) might be generalized here as: “What is true in my and *your* world?”. This “*your*” is indeed a fundamental motivation that underlines our work on coordination in a whole, and the further developments in this paper.

### 3.3 Environments, enumerations, descriptions

We consider the information made available to agents. This information is of two different kinds, and comes from *environments* and *descriptions* as defined below. We assume to have a *full assignment* to all worlds we will consider in the sequel.<sup>2</sup> Our formulation of environments is a restatement of [19] (Definition 3.1A).

---

<sup>2</sup> The notion of *full assignment* we use is standard. For structure  $\mathcal{S}$ , a *full assignment to  $\mathcal{S}$*  is any mapping of *Var onto* the domain of  $\mathcal{S}$ . See for instance [3] for a reference.

**Definition 3.** Let  $\omega$ -sequence  $e$  over  $\mathcal{L}_{basic}$ ,  $\mathcal{L}$ -structure  $\mathcal{S}$ , full assignment  $h$  to  $\mathcal{S}$  and nonempty class of worlds  $\mathbf{K}$  be given.

- (a)  $e$  is a (basic) environment.
- (b)  $e$  is for  $\mathcal{S}$  via  $h$  just in case  $\text{range}(e) = \{\beta \in \mathcal{L}_{basic} \mid \mathcal{S} \models \beta[h]\}$ .
- (c)  $e$  is for  $\mathbf{K}$  just in case  $e$  is an environment for some  $\mathcal{S} \in \mathbf{K}$ .

Thus, an environment is a sequence of increasing, consistent or inconsistent sets of basic formulas. In particular, an environment for  $\mathcal{S}$  (via full assignment  $h$ ) lists the basic diagram of  $\mathcal{S}$  using  $h$  to supply temporary names for the members of  $|\mathcal{S}|$ .<sup>3</sup> Finite initial segments of environments thus recapitulate the information available to a single agent about the underlying structure of evidence at a certain time of observation.

*Enumerations.* When an agent is involved in a coordination game, environments take the form of the (finite, consistent) behaviour of the opponent. We now consider this second kind of information. In contrast to environments, information *by enumeration* is “active” as it comes from interacting agents. Of course, this is possible only for systems with more than one agent. Hence, we suppose there are at least two agents around.<sup>4</sup> Agents should be made able to manage information from different information sources as environments and enumerations. Otherwise, only one-way interaction is possible, that is the interaction between the agent and his “passive” environment. Notice that the information we are looking for does not depend on worlds, but only on the communication abilities of the agents. Thus, next terminology involves basic agents only, leaving worlds-based agents out for further developments.

**Definition 4.** Let agents  $\Psi$  and  $\Phi$  be given.

- (a) The enumeration from  $\Psi$  and  $\Phi$  is the pair  $[\bar{\psi}, \bar{\phi}]$  of  $\omega$ -sequences defined by induction as follows:  $\bar{\psi}_0 = \Psi(\emptyset)$  and  $\bar{\phi}_0 = \Phi(\emptyset)$ . Let  $\bar{\psi}(n) = \langle \bar{\psi}_0 \cdots \bar{\psi}_n \rangle$  and  $\bar{\phi}(n) = \langle \bar{\phi}_0 \cdots \bar{\phi}_n \rangle$ . Then, we define  $\bar{\psi}_{n+1} = \Psi(\bar{\phi}(n))$  and  $\bar{\phi}_{n+1} = \Phi(\bar{\psi}(n))$ .
- (b) Let  $k \in \mathbb{N}$  be given. The enumeration from  $\Psi$  and  $\Phi$  starting at  $k$  is the pair  $[\bar{\psi}^{(k)}, \bar{\phi}^{(k)}]$ , where  $\bar{\psi}^{(k)}$  and  $\bar{\phi}^{(k)}$  are obtained from  $\bar{\psi}$  and  $\bar{\phi}$  by deleting the first  $k + 1$  elements.

The following terminology will also be useful. Let agent  $\Psi$  be given. We say that  $\omega$ -sequence  $\bar{\psi}$  is an *enumeration from  $\Psi$*  just in case  $[\bar{\psi}, \bar{\phi}]$  is the enumeration from  $\Psi$  and  $\Phi$  for some agent  $\Phi$ . We say that  $\bar{\psi}$  is an *enumeration* just in case  $\bar{\psi}$  is an enumeration from  $\Psi$ . We say that  $[\bar{\psi}, \bar{\phi}]$  is the *enumeration from worlds-based agents*  $\langle \Psi_0, \mathbf{A} \rangle$  and  $\langle \Phi_0, \mathbf{B} \rangle$  just in case  $[\bar{\psi}, \bar{\phi}]$  is the enumeration from  $\Psi_0$  and  $\Phi_0$ . To return briefly to the game-theoretic meaning of coordination, let us note that an enumeration  $[\bar{\psi}, \bar{\phi}]$  is *the play* in a coordination game between agents  $\Psi$  and  $\Phi$ . It follows directly from the definition of enumeration that coordination games are *infinite games*. Nevertheless, it is important to observe that *finite* coordination games are possible, and even useful when modeling coordination phenomena within real systems.

<sup>3</sup> We use “basic diagram” as “diagram” in the sense of A. Robinson; see *e.g.* [3].

<sup>4</sup> To simplify matters, we consider here systems of *exactly* two agents.

*Descriptions.* There are several ways in which agents  $\langle \Psi_0, \mathbf{A} \rangle$  and  $\langle \Phi_0, \mathbf{B} \rangle$  may stabilize to a consistent behaviour. To coordinate, agents' successive conjectures must eventually stabilize to a set of formulas that gives them a sufficiently accurate information about one of other agent's worlds. Two concepts must thus be defined: "stabilization" and "sufficient accuracy." Stabilization comes first.

**Definition 5.** *Let  $k \in N$ , agent  $\Psi$  and nonempty class of worlds  $\mathbf{K}$  be given.*

(a)  $\Psi$  ultimately describes  $\mathbf{K}$  just in case enumeration  $\bar{\psi}$  is an environment for some  $S \in \mathbf{K}$ .

(b)  $\Psi$  ultimately describes  $\mathbf{K}$  starting at  $k$  just in case enumeration  $\bar{\psi}^{(k)}$  is an environment for some  $S \in \mathbf{K}$ .

In these cases,  $\bar{\psi}$  and  $\bar{\psi}^{(k)}$  are said to be full descriptions for  $\mathbf{K}$ .

If  $\bar{\psi}$  or  $\bar{\psi}^{(k)}$  is an environment for some world,  $\Psi$  eventually reaches the necessary information to coordinate. Full descriptions are a kind of information made explicit by some agent, and represent the formal, "active" counterpart of the information provided to agents by "passive" environments. However, this information is not sufficient to coordinate, as we see in next section.

### 3.4 Success criteria.

By Definition 5 we give a meaning to "stabilization for an agent in a coordination game": An agent *stabilize* if he or she eventually ends up with an enumeration that is an environment for some world, viz., a full description. How "accurate" such a description have to be to coordinate is the meaning of next definition.

**Definition 6.** *Let  $n, k \in N$ , agents  $\langle \Psi_0, \mathbf{A} \rangle$ ,  $\langle \Phi_0, \mathbf{B} \rangle$  and enumeration  $[\bar{\psi}, \bar{\phi}]$  from  $\langle \Psi_0, \mathbf{A} \rangle$  and  $\langle \Phi_0, \mathbf{B} \rangle$  be given.*

(a)  $\langle \Psi_0, \mathbf{A} \rangle$  cognitively matches with  $\langle \Phi_0, \mathbf{B} \rangle$  at  $n$  just in case  $\bar{\psi}(n)$  is satisfiable in some  $\mathcal{B} \in \mathbf{B}$  and  $\bar{\phi}(n)$  is satisfiable in some  $\mathcal{A} \in \mathbf{A}$ .

(b)  $\langle \Psi_0, \mathbf{A} \rangle$  cognitively matches with  $\langle \Phi_0, \mathbf{B} \rangle$  at  $n$  starting at  $k \leq n$  just in case  $\bar{\psi}^{(k)}(n)$  is satisfiable in some  $\mathcal{B} \in \mathbf{B}$  and  $\bar{\phi}^{(k)}(n)$  is satisfiable in some  $\mathcal{A} \in \mathbf{A}$ .

Observe that  $\mathcal{A}$  and  $\mathcal{B}$  may depend on  $n$ . Moreover, "cognitively matches with" is a reflexive and symmetric binary relation (on  $\Delta$ ). Definition 6 takes two agents to be cognitively matched at some time of the play if there exist two worlds, one for each agent's background world, which satisfy the enumeration provided so far in the play by the other agent. Our conception of coordination simply extends the idea to agents that hold consistency, *i.e.*, agents that ultimately describe some world. What background worlds can provide worlds-based agents to coordinate? To answer that we must first say what we mean by the question. We will distinguish four senses in which two agents could be said to coordinate. In what follows, *f*, *s*, *l* and *mf* may be read as "full," "slow," "local," and "message-failure," respectively.

**Definition 7.** Let agents  $\langle \Psi_0, \mathbf{A} \rangle$ ,  $\langle \Phi_0, \mathbf{B} \rangle$  and enumeration  $[\bar{\psi}, \bar{\phi}]$  from  $\langle \Psi_0, \mathbf{A} \rangle$  and  $\langle \Phi_0, \mathbf{B} \rangle$  be given.

(a) Let  $\mathbf{K}$  be a nonempty class of worlds.  $\langle \Psi_0, \mathbf{A} \rangle$  and  $\langle \Phi_0, \mathbf{B} \rangle$  *lf-coordinate* just in case both  $\langle \Psi_0, \mathbf{A} \rangle$  and  $\langle \Phi_0, \mathbf{B} \rangle$  ultimately describe some  $\mathbf{K}$  and for every  $n \in N$ ,  $\bar{\psi}(n)$  is satisfiable in some  $\mathcal{A} \in \mathbf{A}$ ,  $\bar{\phi}(n)$  is satisfiable in some  $\mathcal{B} \in \mathbf{B}$  and  $\langle \Psi_0, \mathbf{A} \rangle$  cognitively matches with  $\langle \Phi_0, \mathbf{B} \rangle$  at  $n$ .

(b)  $\langle \Psi_0, \mathbf{A} \rangle$  and  $\langle \Phi_0, \mathbf{B} \rangle$  *f-coordinate* just in case  $\langle \Psi_0, \mathbf{A} \rangle$  ultimately describes  $\mathbf{A}$ ,  $\langle \Phi_0, \mathbf{B} \rangle$  ultimately describes  $\mathbf{B}$  and for every  $n \in N$ ,  $\langle \Psi_0, \mathbf{A} \rangle$  cognitively matches with  $\langle \Phi_0, \mathbf{B} \rangle$  at  $n$ .

(c)  $\langle \Psi_0, \mathbf{A} \rangle$  and  $\langle \Phi_0, \mathbf{B} \rangle$  *sf-coordinate* just in case for some  $k \in N$ ,  $\langle \Psi_0, \mathbf{A} \rangle$  ultimately describes  $\mathbf{A}$  starting at  $k$ ,  $\langle \Phi_0, \mathbf{B} \rangle$  ultimately describes  $\mathbf{B}$  starting at  $k$  and for all  $n \geq k$ ,  $\langle \Psi_0, \mathbf{A} \rangle$  cognitively matches with  $\langle \Phi_0, \mathbf{B} \rangle$  at  $n$  starting at  $k$ .

(d) Suppose that for almost all  $i \in N$ ,  $\bar{\psi}_i \neq \perp$  and  $\bar{\phi}_i \neq \perp$ . Let  $k$  be maximal such that either  $\bar{\psi}_k = \perp$  or  $\bar{\phi}_k = \perp$ .  $\langle \Psi_0, \mathbf{A} \rangle$  and  $\langle \Phi_0, \mathbf{B} \rangle$  *mf-coordinate* just in case  $\langle \Psi_0, \mathbf{A} \rangle$  ultimately describes  $\mathbf{A}$  starting at  $k$ ,  $\langle \Phi_0, \mathbf{B} \rangle$  ultimately describes  $\mathbf{B}$  starting at  $k$  and for all  $n \geq k$ ,  $\langle \Psi_0, \mathbf{A} \rangle$  cognitively matches with  $\langle \Phi_0, \mathbf{B} \rangle$  at  $n$ .

## 4 Discussion

“Local full” coordination is the most liberal paradigm, and provides us a model for the more popular problem of coordination between agents. By local coordination agents are allowed to fully describe an arbitrary but common world. In this case, one would say that agents take an *agreement on* that world. The information on the background world that each agent gives to the other agent is thus assured by adding the request for each agent’s output to be finitely consistent with some world in his or her own background world. A question is what classes  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{K}$  allow agents to *lf-coordinate*. Of course, a similar local paradigm may be defined as a generalization of *sf-coordination*. “Full” coordination is a stringent version of local coordination, where each agent must provide a full description of a world in *his* or *her* own background world in place of an arbitrary world taken from the class of all worlds. “Slow full” coordination is a paradigm of coordination *by failure*. Agents are free to stabilize to a suitable description of their worlds after a finite number of failures (disagreements). For this reason, we qualified this paradigm of “slow” coordination. *sf-coordination* arises if agents fail to communicate their clues in the play. According to this paradigm, agents can restart finitely many often, but after the last failure message they must eventually coordinate. The last paradigm, *mf-coordination*, is a particular version of coordination by failure; agents are permitted to explicitly state their coordinating problems by outputting a special atom of their language:  $\perp$ . Again, for agents  $\Psi$  and  $\Phi$ ,  $\bar{\psi}_i$  and  $\bar{\phi}_i$  must differ from  $\perp$  for infinitely many  $i \in N$ , *i.e.* there are only finitely many failure messages. An interesting question on “slow” coordination is how *sf-coordination* compares with *mf-coordination*, and precisely how the use of explicit failure relates with success in any coordination play. It seems likely that something is gained in the efficiency of coordination.

Other definitions of coordination are possible, of course, and an interesting question concerns for what pairs of background worlds what communication abilities provide agents (based on the first and the second element of the pairs, and having that communication abilities) to coordinate. We will not attempt to answer the question save to remark that if the background world of both agents is a singleton,  $f$ -coordination models a situation rather close to that of Ehrenfeucht games for elementary equivalence in model theory (see for instance [4] for background on Ehrenfeucht games). It seems likely that  $f$ -coordination is equivalent to the interactive construction of partial isomorphisms between the worlds of the agents involved, but we have no proofs to give.

There are many potentially interesting generalizations of the logical framework just described:

1. Environments of the kind introduced here are *basic environments*, in the sense their expressive power goes along with basic formulas. However, it is sometimes useful to have more expressive environments. An agent's observation may require more information than is captured in basic formulas. Thus, it could be both necessary and useful to extend environments over quantified formulas. Of course, this enrichment rises new problems when extending the resulting framework towards computational settings, our next point.

2. One can develop the conceptual framework into computational issues. A restriction to computer simulable agents as well to *finite* coordination games is required to study the coordination problem within real multi-agent systems. Computable agents can be obtained by considering agents based on recursive classes of finite worlds to recursive, or possibly P-TIME communication abilities. Agents might be bounded in their computation time, *i.e.* as a function of the length of the input. The speed of coordination may be taken into account. Thus, given worlds-based agent  $\Psi$  and  $n \in \mathbb{N}$ , one can define the collection of worlds-based agents  $\Phi$  such that  $\Psi$  changes his actual world (coordinating with  $\Phi$  according to each paradigm) no more than  $n$  times (cf. [9] for some reference on "mind changes" in inductive inference).

3. So far, no mention has been given to how coordination can be used to problem solving. By extending the agents' outputs with a second component ranging over arbitrary sets of formulas, coordination is suitable to investigate *Team*-solvability. Coordination moves to *cooperation*. Then the question is: What classes of worlds are solvable (according to some paradigm of inquiry; we refer the reader to [13, 12, 16]) by a *team* of agents? A team of  $m \geq 3$  agents can be defined as a set of pair of agents that coordinate according to some paradigm of coordination. A similar approach to team solvability—called *Team*-identification, has been developed within *Formal Learning Theory* (see [9], Chapter 9). However, a definition of *Team*-identifiability is given for total recursive functions (Smith, [26]) and recursive enumerable formal languages (Osherson *et. al.*, [17]). As far as we know, no *Team*-identifiability is given for first-order structures. More important, definitions by Smith and Osherson *et. al.* "fail to formalize one aspect of scientific practice that is central to the informal idea of team work." ([9], Chapter 9, pp.198-199). "The hypotheses of the individual scientists [agents] in

team scientific discovery influence each other in a way that is not captured by Smith and Osherson *et. al.* definitions.” The paradigms of coordination are a suitable starting point to capture the hint of the informal idea of teamwork we look at, and provide the “formation rule” for teams of collaborative agents.

4. The problem of belief change—how an agent should revise her beliefs upon learning new information—can be taken into account. One can investigate belief changes by limiting a slightly modified version of the paradigms of coordination to agents based on background worlds that are expressible by a finite set of basic formulas. For such finite sets, one can put *revision* into the communicative ability  $\Psi_0$  of agent  $\langle \Psi_0, \mathbf{A} \rangle$  by assuming  $\Psi_0 = \lambda\sigma.K \# \sigma$ , where  $\# : \text{pow}(\mathcal{L}_{\text{basic}}) \times \text{SEQ} \rightarrow \text{pow}(\mathcal{L}_{\text{basic}})$  is a suitable revision function and  $\mathbf{A} = \text{MOD}(K)$ . One can take  $\#$  to be “rational” according to some principle of rationality (see for instance [6, 10, 13, 12] and the reference listed there). Agents of this new sort have “rational” communication abilities in this strict sense. It is then possible to investigate the coordination game of  $\langle \lambda\sigma.A \#_a \sigma, \mathbf{A} \rangle$  and  $\langle \lambda\sigma.B \#_b \sigma, \mathbf{B} \rangle$  into three winning strategies: (a) strategies that hold constant the background *belief sets*  $A$  and  $B$ , and then determine what kind of revision functions allow agents to coordinate; (b) strategies that hold constant the revision functions and determine what kind of belief sets allow agents to coordinate; (c) strategies that hold constant both belief sets and revision functions, and determine what kind of coordination paradigm allow agents to coordinate.

In the context of related work, there is a previous framework for analyzing the problem of coordination in the limit. That account, due to Franco Montagna and Daniel Osherson [15], is in the spirit of the recursion-functional tradition of inductive inference in something like the way that the framework given here is in the spirit of model-theoretic’s. In Montagna and Osherson’s work, the agent’s communication skills are investigated by defining several kinds of players that interactively “learn to coordinate”. Montagna and Osherson take the coordination problem of two agents or *players* that want to coordinate by repeatedly showing each other one of two possible behaviours. The problem of coordination the players are faced with follows from the shifting constraints on their behaviours. Each player tries to predict the other’s behaviour, and their predictions are based on no more than the history of earlier events. One player “learns” the other’s behaviour if her or his own behaviour matches the other’s forever after. There is no an unique winner in the coordination game. To keep matters simple, Montagna and Osherson consider two players facing the same two options on each trial, and they denote the options by 0 and 1. A player is therefore be identified with a 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. There are several obvious differences in the frameworks, some of which being a direct consequence of the language in use (recursive-theoretic vs. first-order), and one that is not obvious but which is the most important: *sf*-coordination extends the learning to coordinate Montagna and Osherson’s paradigm, in the sense that coordination arises between Montagna and Osherson’s players if and only if *sf*-coordination arises between a special kind of agents defined on them.

## References

1. L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28(2):125–155, 1975.
2. P. Brazdil, M. Gams, S. Sian, L. Torgo, and W. van de Velde. Learning in distributed systems and multi-agent environments. In Y. Kodratoff, editor, *Machine Learning - European Working Session on Learning*, pages 412–423. Springer-Verlag LNAI 482, 1991.
3. C.C. Chang and J.M. Keisler. *Model Theory - 3rd edition*. North Holland, 1990.
4. H-D Ebbinghaus and J. Flum. *Finite Model Theory*. Springer, 1995.
5. S. Franklin and A. Graesser. Is it an agent, or just a program?: A taxonomy for autonomous agents. In J. P. Müller, M. J. Wooldridge, and N. R. Jennings, editors, *Intelligent Agents III - Agent Theories, Architectures, and Languages*, pages 21–35. Springer-Verlag LNAI 1193, 1997.
6. P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press, Cambridge, MA, 1988.
7. C. Glymour. Inductive inference in the limit. *Erkenntnis*, 22:23–31, 1985.
8. E. M. Gold. Language identification in the limit. *Information and Control*, 10:447–474, 1967.
9. 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. MIT Press, Cambridge, MA, 1999.
10. K. T. Kelly. *The Logic of Reliable Inquiry*. Oxford University Press, New York, NY, 1996.
11. D. K. Lewis. *Conventions. A Philosophical Study*. Harvard University Press, Cambridge, MA, 1969.
12. E. Martin and D. Osherson. Scientific discovery based on belief revision. *Journal of Symbolic Logic*, 62(4):1352–1370, 1997.
13. E. Martin and D. Osherson. *Elements of Scientific Inquiry*. MIT Press, Cambridge, MA, 1998.
14. E. Mendelson. *Introduction to Mathematical Logic - 3rd edition*. The Wadsworth & Brooks/Cole mathematics series. Wadsworth, Monterey, CA, 1987.
15. F. Montagna and D. Osherson. Learning to coordinate: A recursion theoretic perspective. *Synthese*, in press.
16. 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 Publishers B.V., 1997.
17. D. Osherson, M. Stob, and S. Weinstein. *Systems That Learn*. The MIT Series in Learning, Development, and Conceptual Change, v. 4. MIT Press, Cambridge, MA, 1986.
18. D. Osherson, M. Stob, and S. Weinstein. A universal inductive inference machine. *Journal of Symbolic Logic*, 56(2):661–672, 1991.
19. D. Osherson and S. Weinstein. Identification in the limit of first order structures. *Journal of Philosophical Logic*, 15:55–81, 1986.
20. E. Plaza, J. L. Arcos, and F. Martín. Cooperative case-based reasoning. In G. Weiß, editor, *Distributed Artificial Intelligence meets Machine Learning*, pages 180–201. Springer-Verlag LNAI 1221, 1997.
21. H. Putnam. Trial and error predicates and a solution to a problem of Mostowski. *Journal of Symbolic Logic*, 30(1):49–57, 1965.

22. S. Sen, M. Sekaran, and J. Hale. Learning to coordinate without sharing information. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, pages 426–431, Seattle, 1994.
23. E. Shapiro. Inductive inference of theories from facts. In J-L. Lassez and G. Plotkin, editors, *Computational Logic: Essays in honor of Alan Robinson*. MIT Press, 1991.
24. Y. Shoham and M. Tennenholtz. Emergent conventions in multi-agents systems: Initial experimental results and observations. In *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*, pages 225–231, Cambridge, 1992.
25. Y. Shoham and M. Tennenholtz. On the synthesis of useful social laws for artificial agent societiesp. In *Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 276–281, San Jose, 1992.
26. C. Smith. The power of pluralism for automatic program synthesis. *Journal of the ACM*, 29:1144–1165, 1982.
27. R. J. Solomonoff. A formal theory of inductive inference. *Information and Control*, 7:7–22, 1964.
28. G. Weiß. Learning to coordinate actions in multi-agent systems. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 311–316, Chambéry, France, 1993.
29. G. Weiß. Adaptation and learning in multi-agents systems: Some remarks and a bibliography. In G. Weiß and S. Sen, editors, *Adaptation and Learning in Multi-Agent Systems*, pages 1–21. Springer-Verlag LNAI 1042, 1995.
30. M. J. Wooldridge and N. R. Jennings. Agent theories, architectures, and languages: A survey. In M. J. Wooldridge and N. R. Jennings, editors, *Intelligent Agents*, pages 1–39. Springer-Verlag LNAI 890, 1995.