Team-solvability: A Model-Theoretic Perspective

Alessandro Agostini[1]

Abstract. At present, the extension of *formal learning theory* to the multi-agent case considers "teams" of agents sharing a common end. Success is achieved if one or more of the agents is successful, and cooperation is not involved in the team formation. Unfortunately, this is rarely the idea of "successful team" we have in mind. One generally expects agents' behavior to influence each other in a way that is not captured by the present paradigms. A real problem in extending single agent learning methods to multi-agent setting is thus determining *paradigms of cooperation*. This paper makes a contribution to the solution of this problem. First, we advance a paradigm of cooperation as a kind of two-person repeated game and compare it to a major paradigm of solvability for isolated agents. Second, we pay attention to a subset of *unsuccessful* agents who take advantage from teamwork. For these agents, cooperation is proved to be a key of success. The formal results are raised within the model-theoretic tradition of formal learning theory.

1 INTRODUCTION

There has been a growing interest in AI in the design and theorization of systems of multiple autonomous agents that interact in various ways as they pursue their own ends, or perhaps seek compatible goals. Of special interest are systems in which individual agents share the same goals or utility function. In such settings, the agents have to choose between two course of actions. Thus, the agents either act collectively or separately to the common desired ends. Two collective actions are more investigated than others: coordination and cooperation. The former has received the bulk of attention by AI (*e.g.*, [4, 22, 26, 1]) and game theory, where the coordination space arising from a successful coordination problem is called *equilibrium* (see for instance [15, 14]). The latter has been extensively studied in AI, where fully cooperative problems arise in task distribution as well as within the historically older sub-area of the multi-agent problem solving (*e.g.*, [20, 28, 29]).

In this paper, our goal is to investigate cooperation and teamwork from a *learning-by-discovery* [27] perspective. To this end, we focus on the model-theoretic tradition of *Formal Learning Theory*—say [23, 7, 18, 17, 11], that descends from the pioneering studies on inductive inference developed by [25, 21, 8, 3]. 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 by Montagna and Osherson [13] to characterize recursive functions by means of coordination in the limit. Coordination has been recently addressed into the model-theoretic tradition in [1, 2]. Since learnability and problem solving have been studied extensively in formal learning theory and AI in the context of single agent systems, the question naturally arises: to what extent can models or *paradigms* for single agent learning be extended to the cooperative multi-agent setting?

The paper is thus structured as follows. In Section 2 we briefly review the paradigm of $\mathbf{Ex}^{\pi}$-solvability [16], our starting point. We also introduce preliminary concepts and notation that will be used throughout the whole paper. Section 3 presents the main contribution of the paper. It divides into two related parts, involving respectively cooperative and unsuccessful agents problem solving. Finally, in Section 4 we add some further remarks on related and future work and the conclusion.

2 PRELIMINARY CONCEPTS AND NOTATION

We fix a 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 (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 $\{0, 1, 2, ...\}$ of natural numbers is denoted by N, the set $\{1, 2, ...\}$ of positive natural numbers is denoted by N^+. If X is a set, X^ω is the set of infinite sequences over X. A sequence in X^ω is called an ω*-sequence* (over X). Let ϱ be an ω-sequence. We write $\varrho(i)$, $i \in N$, for the finite sequence $\langle \varrho_0 \cdots \varrho_i \rangle$, and $\varrho|_i$ for the proper initial segment of length i in ϱ. Thus, $\varrho(i) = \varrho|_{i+1}$. We write *length*$(\eta)$ for the length of a finite sequence and η_i for the ith element of η, $0 \leq i <$ *length*(η). We write *range*(η) for the set of elements of any sequence. We denote the finite sequence of length zero by $\emptyset$. Otherwise, our semantic notions are standard. In particular, structure $\mathcal{S}$ is a model of $\Gamma \subseteq \mathcal{L}_{form}$, and Γ is said to be *satisfiable in* $\mathcal{S}$, if there is an assignment $h : Var \longrightarrow |\mathcal{S}|$ with $\mathcal{S} \models \Gamma[h]$. Γ is *satisfiable* if it is satisfiable in some structure.

2.1 Games I: Isolated agents against Nature

The following picture of scientific inquiry cited from [16] gives an informal idea of the basic elements of the game we are interested in. "First, a class of possible realities is specified in advance; the class is known to both players of the game. Nature is conceived as choosing one member from the class to be the 'actual world'; her choice

[1] LOMIT - Dipartimento di Matematica, Università di Siena, via del Capitano 15, 53100 Siena, Italy; email: agostini@unisi.it

© 2000 A. Agostini
ECAI 2000. *14th European Conference on Artificial Intelligence*
Edited by Werner Horn
Published in 2000 by IOS Press

is initially unknown to the scientist [agent]. Nature then provides a series of clues about this reality. These clues constitute the data upon which the scientist will base his hypotheses. Each time Nature provides a new clue, the scientist may produce a new hypothesis. The scientist wins the game if there is sufficient guarantee that his successive conjectures will stabilize to an accurate hypothesis about the reality Nature has chosen." (p. 740)

We mention possible realities and "worlds". Formally, by *world* we shall mean any countable structure that interprets $\mathcal{L}$, or *$\mathcal{L}$-structure*. Worlds may be conceived as the "possible truths" for the agents. We shall be 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". In the sequel, we shall use "structure" in place of "$\mathcal{L}$-structure".

Definition 1 *Let mapping Ψ' : SEQ $\longrightarrow$ $\mathcal{L}_{sen}$ and nonempty class* **A** *of structures be given. We say that $\Psi = \langle \Psi', \mathbf{A} \rangle$ is a (basic)* agent.

For all $\sigma \in$ *SEQ*, we then write $\Psi(\sigma)$ for $\Psi'(\sigma)$. We say that $\mathcal{L}_{sen}$ is the *agent's language*. According to the terminology adopted in the literature within formal learning theory, if **A** is empty we say that $\langle \Psi', \emptyset \rangle$ is a (basic) *scientist* or also a *learner*. Agent might be partial or total, recursive or nonrecursive. Although we will keep our discussion as general as possible, particular attention to computational agents is given in multi-agent systems. Thus, we can assume to deal with computable agents. Of the two components of any agent, the first is said to be the *agent's communication ability* and the second component is said to be the *agent's background world*. To fix intuitions one might think of a background world as representing the agent's belief space. In the picture of scientific inquiry above, some further elements need to be explained.

We consider the information made available to agents. This information is of two different kinds, and comes from "environments" as defined below. We assume to have an *assignment* to all worlds we will consider in the sequel.[2] Our formulation of environments is a restatement of [18] (Definition 3.1A).

Definition 2 *Let e be an ω-sequence over $\mathcal{L}_{basic}$. We say that e is a* (basic) environment*. Let world $\mathcal{S}$ and full assignment h to $\mathcal{S}$ be given. We say that environment e is* for $\mathcal{S}$ via h *just in case range$(e) = \{\beta \in \mathcal{L}_{basic} \,|\, \mathcal{S} \models \beta[h]\}$.*

Thus, an environment is a sequence of increasing, consistent or inconsistent sets of basic formulas. In particular, an environment for $\mathcal{S}$ (via assignment h) lists the basic diagram of $\mathcal{S}$ using h to supply temporary names for the members of $|\mathcal{S}|$.[3] 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.

Definition 3 *Let $\theta \in \mathcal{L}_{sen}$, environment e and agent Ψ be given. Ψ* converges on e to θ *just in case for cofinitely many $k \in N$, $\Psi(e|_k) = \theta$.*

The following definition is on "approximate" solvability (see *e.g.* [19] for a discussion on approximate solvability). The next definition is a restatement of [16] (Definition 27 and 29).

Definition 4 *Let $\pi \subseteq \mathcal{L}_{sen}$, agent Ψ, structure $\mathcal{S}$ and environment e be given. Suppose that e is for $\mathcal{S}$.*

1. Ψ $\mathbf{Ex}^{\pi}$-solves e *just in case there is $\theta \in \pi$ such that Ψ converges on e to θ and $\mathcal{S} \models \theta$.*
2. Ψ $\mathbf{Ex}^{\pi}$-solves $\mathcal{S}$ *just in case Ψ $\mathbf{Ex}^{\pi}$-solves every environment for $\mathcal{S}$.*
3. Ψ $\mathbf{Ex}^{\pi}$-solves collection **K** of structures *just in case Ψ $\mathbf{Ex}^{\pi}$-solves every $\mathcal{S} \in$ **K**. In this case,* **K** *is said to be* $\mathbf{Ex}^{\pi}$-solvable.
4. $\mathbf{Ex}^{\pi}$ *is the collection of* $\mathbf{Ex}^{\pi}$*-solvable classes of structures.*

Definition 4 completes the formalization of the elements that figure in the game-theoretic picture of scientific inquiry. Two further remarks and an interesting question arise. First, agents do not use their background world. This fact is fairly close to the general conception of learning as empirical inquiry [12, 9]. An agent *could* use her background world in principle; for example, a belief-revision based agent could eventually represent some "belief state" by using her background world. So, the problem of belief change—how an agent should revise her beliefs upon learning new information, can be taken into account. For lack of space, we do not discuss this topic here. Second, the paradigm of $\mathbf{Ex}^{\pi}$-solvability is exactly the model of *X-solvability* given in the literature for "approximate solvability" (see for instance [16, 19]). Third, the question: What happens if many agents jointly work to a problem?

3 MULTI-AGENT PROBLEM SOLVING

Consider the case of the isolated agent facing nature: Time and resources are scarce, and there may be risk or uncertainty about future states of the world. $\mathbf{Ex}^{\pi}$-solvability theory tell us how such an agent will decide when facing different circumstances: He has preferences and beliefs and is rational according to some principle of rationality (see for instance [6, 10, 12, 11] and the reference listed there).

Suppose now we introduce other agents into out agent's environment and make them interact. Is a theory of their interaction reducible to a theory of the isolated agent? One might wonder why there should be any difficulty here. After all, the only difference between a natural environment and a social environment is just the presence of other people; rational choice looks the same in both cases. To answer that we must first say what we mean by "interaction". In what follows we consider a very special kind of interaction, ie: *cooperation*. Indeed, cooperation allow us to shift the single-agent processes of the paradigm $\mathbf{Ex}^{\pi}$ into actual teamwork, in the sense that convergence to a stable state (partial solution) is achieved only if a cooperative response is given by the agents on the common, possibly partial representation of the problem to be solved.

3.1 Games II: Cooperative agents against Nature

To address the model or *paradigm* of cooperation formally, we focus on the importance of communication in cooperative actions w.r.t., say, mental attitude. Thus, we extend agents to "collaborative" agents as follows. (For any set X, let *pow*(X) denote the power set of X.)

Definition 5 *Let mapping Ψ' from pow$(\mathcal{L}_{sen}) \times$ SEQ to $\mathcal{L}_{basic} \times \mathcal{L}_{sen}$ and nonempty class* **A** *of structures be given. We say that $\Psi = \langle \Psi', \mathbf{A} \rangle$ is a* collaborative agent*.*

We say that $\mathcal{L}_{basic} \cup \mathcal{L}_{sen}$ is the agent's language. Similarly to basic agents, collaborative agents may be partial or total, computable

[2] The notion of *assignment* we use is standard. For structure $\mathcal{S}$, an *assignment to $\mathcal{S}$* is any mapping of *Var onto* the domain of $\mathcal{S}$. See for instance [5] for a reference.

[3] We use "basic diagram" as "diagram" in the sense of A. Robinson; see *e.g.* [5].

or noncomputable. For all $\sigma \in SEQ$ and all $\Theta \subseteq \mathcal{L}_{sen}$, we write $\Psi(\Theta, \sigma)$ for $\Psi'(\Theta, \sigma)$. Moreover, observe that $\Psi(\Theta, \sigma) = \langle (\Psi(\Theta, \sigma))_0, (\Psi(\Theta, \sigma))_1 \rangle$. To help intuitions, for every agent Ψ's input Θ,σ, one might think to $(\Psi(\Theta, \sigma))_0$ as the "public output" of the agent. This component is used by the agent to communicate with other agents; it is, say, a *social* component. In contrast, $(\Psi(\Theta, \sigma))_1$ may be interpreted as the "private output" of the agent. This second component is used by the agent to guess solutions and doing hypotheses in problem solving. We note that for fixed $\Theta \subseteq \mathcal{L}_{sen}$, $\lambda\sigma.(\Psi(\Theta, \sigma))_1$ is a basic agent. Thus, $\lambda\sigma.(\Psi(\Theta, \sigma))_1$ is a kind of "oracle" that the agent provides to a second agent interacting with him in a inquiry process.[4] The agents private output is what makes collaborative agents similar to basic, say *noncollaborative*, agents. The usage meaning of the second component is made close to basic agents' behavior in the next definition, which thus generalizes Definition 3 to collaborative agents. Let $\theta \in \mathcal{L}_{sen}$, environment e and collaborative agent Ψ be given. We say that Ψ *converges on e to θ* just in case for some $\Theta \subseteq \mathcal{L}_{sen}$, $\lambda\sigma.(\Psi(\Theta, \sigma))_1$ converges on e to θ. Suppose that environment e is for some world made actual by Nature. Then, the definition makes clear what we mean by a collaborative agent communicating with Nature: Definition 4 extends onto collaborative agents in the obvious way. It remains to see in what sense collaborative and noncollaborative agents differ in their matches against Nature.

3.1.1 Cooperation

In a multi-agent setting, information seems to be coming from essentially two quite different sources: Nature and agents. When an agent interacts with an agent, an environment is often the behavior of the opponent. Roughly, we call this behavior *enumeration*. In contrast to environments, information from enumerations is thus "active". 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. The information we are looking for does not depend on worlds, but only on the agents' communication abilities. At this stage of development, this fact reflects an "external" [24], say *communicative* perspective on cooperative activity and team formation. In the sequel we shall see how this external perspective combines with an "internal" perspective. For now, we only record that next terminology does not involve worlds. Let environment e and collaborative agents Ψ and Φ be given. The *enumeration from Ψ and Φ in e* is the pair $[\overline{\psi}(e) \, , \overline{\phi}(e)]$ of pairs of ω-sequences $\overline{\psi}(e) = \langle \overline{\psi}_0, \overline{\psi}_1 \rangle$ and $\overline{\phi}(e) = \langle \overline{\phi}_0, \overline{\phi}_1 \rangle$ defined by induction as follows. We define $\overline{\psi}_{00} = (\Psi(\emptyset, \emptyset))_0$ and $\overline{\psi}_{01} = (\Psi(\emptyset, \emptyset))_1$; $\overline{\phi}_{00} = (\Phi(\emptyset, \emptyset))_0$ and $\overline{\phi}_{01} = (\Phi(\emptyset, \emptyset))_1$. Let $\overline{\psi}_1|_n = \langle \overline{\psi}_{01} \cdots \overline{\psi}_{(n-1)1} \rangle$ and $\overline{\phi}_1|_n = \langle \overline{\phi}_{01} \cdots \overline{\phi}_{(n-1)1} \rangle$. Then, we define $\overline{\psi}_{0n} = (\Psi(\mathit{range}(\overline{\phi}_1|_n), e|_n))_0$ and $\overline{\phi}_{0n} = (\Phi(\mathit{range}(\overline{\psi}_1|_n), e|_n))_0$; and $\overline{\psi}_{1n} = (\Psi(\mathit{range}(\overline{\phi}_1|_n), e|_n))_1$ and $\overline{\phi}_{1n} = (\Phi(\mathit{range}(\overline{\psi}_1|_n), e|_n))_1$. Let $k \in N$ be given. The *enumeration from Ψ and Φ in e starting at k* is the pair $[\overline{\psi}(e)^{(k)} \, , \overline{\phi}(e)^{(k)}]$, where $\overline{\psi}(e)^{(k)}$ and $\overline{\phi}(e)^{(k)}$ are obtained from $\overline{\psi}(e)$ and $\overline{\phi}(e)$ by deleting the first $k+1$ elements in $\overline{\psi}_0$ and $\overline{\phi}_0$, respectively. In the rest of this paper, we sometimes write $R\langle \Psi_0, \Phi \rangle$ for $\overline{\psi}_0$ and k-$R\langle \Psi_0, \Phi \rangle$ for $\overline{\psi}_0{}^{(k)}$. In both the notation, we shall leave implicit the environment the enumeration from Ψ and Φ is in. Thus, $R\langle \Psi_0, \Phi \rangle$ is meant as Ψ's *public response to Φ in* some environment.

Our interest is in *two-person cooperation games*. These are defined by two agents whose play is uniquely defined by their communication abilities ("external" perspective, cf. [24]). The agents have internal states to serve them as a basis of choices (*e.g.*, beliefs and expectations; cf. the "internal" perspective, *ibidem*). An intuitive conception of cooperation implies that agents have the same outcomes in principle. Following this conception, we formally introduce cooperation as follows.

Definition 6 *Let $\theta \in \mathcal{L}_{sen}$, environment e, collaborative agents $\langle \Psi, \mathbf{A} \rangle$ and $\langle \Phi, \mathbf{B} \rangle$ and enumeration $[\overline{\psi}(e) \, , \overline{\phi}(e)]$ from $\langle \Psi, \mathbf{A} \rangle$ and $\langle \Phi, \mathbf{B} \rangle$ in e be given.* $\langle \Psi, \mathbf{A} \rangle$ and $\langle \Phi, \mathbf{B} \rangle$ cooperate in e with respect to θ *just in case for some $k \in N$, $\overline{\psi}_0{}^{(k)}$ is an environment for some $\mathcal{A} \in \mathbf{A}$, $\overline{\phi}_0{}^{(k)}$ is an environment for some $\mathcal{B} \in \mathbf{B}$, for cofinitely many $n \in N$, $\overline{\psi}_{1n} = \overline{\phi}_{1n} = \theta$, $\mathcal{A} \models \theta$ and $\mathcal{B} \models \theta$. In this case, we call θ a* cooperation sentence *and we say that $\langle \Psi, \mathbf{A} \rangle$ and $\langle \Phi, \mathbf{B} \rangle$ are* cooperative.

The next definition fixes the criterion of success for cooperative agents problem solving.

Definition 7 *Let $\pi \subseteq \mathcal{L}_{sen}$, collaborative agents Ψ and Φ, structure $\mathcal{S}$ and environment e be given. Suppose that e is for $\mathcal{S}$.*

1. Ψ $\mathbf{Co}^{\pi}$-solves e with Φ *just in case there is $\theta \in \pi$ such that Ψ and Φ cooperate in e with respect to θ and $\mathcal{S} \models \theta$.*
2. Ψ $\mathbf{Co}^{\pi}$-solves $\mathcal{S}$ with Φ *just in case Ψ $\mathbf{Co}^{\pi}$-solves every environment for $\mathcal{S}$ with Φ.*
3. *Let $\mathbf{K}$ be a class of structures.* Ψ $\mathbf{Co}^{\pi}$-solves $\mathbf{K}$ with Φ *just in case Ψ $\mathbf{Co}^{\pi}$-solves every $\mathcal{S} \in \mathbf{K}$ with Φ. In this case, we say that $\mathbf{K}$ is $\mathbf{Co}^{\pi}$-solvable with Φ.*
4. $\mathbf{Co}^{\pi} = \{\mathbf{K} \mid \Psi$ $\mathbf{Co}^{\pi}$*-solves $\mathbf{K}$ with Φ for some Ψ and Φ*$\}$.

With Definition 7 in hand it is easy to prove a result on the limit of cooperation as a paradigm of problem solving.

Proposition 1 *Let $\pi \subseteq \mathcal{L}_{sen}$.* $\mathbf{Co}^{\pi} = \mathbf{Ex}^{\pi}$.

The result may be viewed as a fundamental limitation to use of cooperation to enlarge the collection of solvable problems (according to the paradigm $\mathbf{Ex}^{\pi}$). However, it could be discussed a sense in which data (represented by an environment for some "actual" world) can be used more efficiently by cooperating with some agent that without.[5]

3.2 Games III: Unsuccessful agents against Nature

In this section we study a paradigm of problem solving where cooperative behavior of some "unsuccessful" agents is proved to be useful to improve the agents problem solving ability. Our aim is therefore to give some further insight to the paradigm $\mathbf{Co}^{\pi}$ and its connections with $\mathbf{Ex}^{\pi}$-solvability.

3.2.1 Unsuccessful agents

Let agent Ψ be unsuccessful. Informally, this means that there is at least a structure $\mathcal{S}$ that Ψ cannot solve. However, is often the case that an agent that does not solve a problem for some reason, solves some part of the problem. Then, a suitable definition of "unsuccessful" agent should take into account both ability and limits of the agent. Formally, let collaborative agent Ψ and structure $\mathcal{S}$ be given.

[4] Cf. the definition and use of oracles in empirical inquiry as stated *e.g.* in Martin and Osherson's book [12], Section 3.4.3.

[5] We do not discuss efficiency of cooperative problem solving here.

We say that Ψ is $\mathbf{Ex}^{\pi}$*-unsuccessful on* $\mathcal{S}$ just in case Ψ does not $\mathbf{Ex}^{\pi}$-solve a finite, positive number of environments for $\mathcal{S}$ and Ψ $\mathbf{Ex}^{\pi}$-solves some environment for $\mathcal{S}$. Now, suppose that structure $\mathcal{S}$ is $\mathbf{Ex}^{\pi}$-solvable. Then Ψ can try to solve $\mathcal{S}$ by looking for cooperation. We do not investigate here the general reasons that lead Ψ to recognize the "potential of cooperative action" (cf. [29], p. 574, where this first stage of the cooperative problem-solving process is called: *Recognition*). In our setting, these reasons are two, namely: for some environment e for $\mathcal{S}$, either Ψ does not converge on e to any sentence or Ψ converges on e to some sentence which is false in $\mathcal{S}$. Despite of recognition arises "because [(a)] an agent has a goal that it does not have the ability to achieve on its own or else because [(b)] the agent prefers a cooperative solution" (*ibidem*, p. 574), our paradigm formalizes (a) and does not formalize (b).[6]

3.2.2 Teamwork

How could Ψ solve $\mathcal{S}$ by teamwork? In previous section we saw that Ψ solves at least a part of $\mathcal{S}$, namely, he solves some environment for it. A *conditio sine qua non* for Ψ to be assisted is to communicate the problem. Of course, next step for Ψ shall be to ensure that the agent who response to his request is eventually able to help. This is roughly the meaning of *team formation* in [29]. We formalize in our framework the fact that Ψ recognizes the potential of a cooperative action by forcing Ψ to be consistent on $\mathcal{S}$ in the following sense.

Definition 8 *Let collaborative agent* Ψ *and nonempty class* $\mathbf{A}$ *of structures be given.* Ψ *is* $\mathbf{A}$-consistent *just in case for every collaborative agent* Φ*,* $R\langle\Psi_0,\Phi\rangle$ *is an environment for some* $\mathcal{A}\in\mathbf{A}$.

To remind the story, Ψ does not solves $\mathcal{S}$, and asks for help. Then, Definition 8 says that a necessary condition for Ψ to find some agent that eventually helps him is to be $\{\mathcal{S}\}$-consistent. It follows that there must be in Ψ's background world a structure elementarily equivalent to $\mathcal{S}$.[7] In other words, Ψ must be able to communicate consistent and complete information on $\mathcal{S}$ which is *potentially* known to the agent. We emphasize "potentially" because is not assumed in all our paradigms that agents are aware of their worlds. The order by which Ψ communicates the information will depend on the helper who responds to Ψ. Our requirement on Ψ to form a team is not sufficient. To see why, we need a new definition of teamwork success.

Definition 9 *Let* $\pi\subseteq\mathcal{L}_{sen}$*,* $m\in N^{+}$*, collaborative agent* Ψ*, structure* $\mathcal{S}$ *and environment* e *be given. Suppose that environment* e *is for* $\mathcal{S}$.

1. Ψ $\mathbf{Co}^{m}\mathbf{Ex}^{\pi}$-solves e *just in case there is a set* $\Pi=\Pi(e)$ *of* m *agents such that for every* $\Phi\in\Pi$,
   - *i. there is* $k\in N$ *such that* $k\text{-}R\langle\Psi_0,\Phi\rangle=e$*;*
   - *ii.* Φ *converges on* e *to some* $\theta\in\pi$ *with* $\mathcal{S}\models\theta$*, and*
   - *iii.* Ψ *and* Φ *cooperate in* e *with respect to* θ.
2. Ψ $\mathbf{Co}^{m}\mathbf{Ex}^{\pi}$-solves $\mathcal{S}$ *just in case* Ψ $\mathbf{Co}^{m}\mathbf{Ex}^{\pi}$*-solves every environment for* $\mathcal{S}$.
3. Ψ $\mathbf{Co}^{m}\mathbf{Ex}^{\pi}$-solves class $\mathbf{K}$ of structures *just in case* Ψ $\mathbf{Co}^{m}\mathbf{Ex}^{\pi}$*-solves every* $\mathcal{S}\in\mathbf{K}$*. In this case, we say that* $\mathbf{K}$ *is* $\mathbf{Co}[\Psi]^{m}\mathbf{Ex}^{\pi}$*-solvable.*
4. $\mathbf{Co}[\Psi]^{m}\mathbf{Ex}^{\pi}=\{\mathbf{K}\mid\Psi\ \mathbf{Co}^{m}\mathbf{Ex}^{\pi}\text{-}solves\ \mathbf{K}\}$.

We say that the set $\Pi(e)$ is *helpful for agent* Ψ *in environment* e. Each agent in the set succeeds in helping Ψ in solving a part of the problem. More generally, we have:

Proposition 2 *Suppose* $\mathbf{K}$ *be a* $\mathbf{Ex}^{\pi}$*-solvable class of structures. For all collaborative agents* Ψ*, if* Ψ *is* $\mathbf{Ex}^{\pi}$*-unsuccessful on some* $\mathcal{S}\in\mathbf{K}$ *and* $\{\mathcal{S}\}$*-consistent, then there is* $m\in N^{+}$ *such that* $\mathbf{K}\in\mathbf{Co}[\Psi]^{m}\mathbf{Ex}^{\pi}$.

Proof: (sketched) Let $\mathcal{S}\in\mathbf{K}$ and $\{\mathcal{S}\}$-consistent collaborative agent Ψ be given. Suppose that Ψ does not $\mathbf{Ex}^{\pi}$-solve environment e_1 for $\mathcal{S}$ and that Ψ $\mathbf{Ex}^{\pi}$-solves environment e for $\mathcal{S}$. We need to show that there is $m\in N^{+}$ such that $\mathbf{K}\in\mathbf{Co}[\Psi]^{m}\mathbf{Ex}^{\pi}$. To this end, we define collaborative agent Φ^1 such that: (a) for some $k\in N$, $k\text{-}R\langle\Psi_0,\Phi^1\rangle=e_1$ and $k\text{-}R\langle\Phi^1_0,\Psi\rangle=e$; (b) Φ^1 converges on e_1 to some $\theta\in\pi$ such that $\mathcal{S}\models\theta$; (c) Ψ and Φ^1 cooperate in e_1 with respect to θ. Since $\mathcal{S}$ is $\mathbf{Ex}^{\pi}$-solvable, Ψ is $\{\mathcal{S}\}$-consistent by assumption and $R\langle,\rangle$ is uniquely defined on e_1, it is easy to verify that such Φ^1 exists. Similarly, for every environment e_i for $\mathcal{S}$ such that Ψ does not $\mathbf{Ex}^{\pi}$-solve e_i, we define agent Φ^i. Because of Ψ is $\mathbf{Ex}^{\pi}$-unsuccessful on $\mathcal{S}$ by assumption, it follows that the resulting set of agents is finite and nonempty. Let $\{\Phi^1,\ldots,\Phi^m\}$ such finite set with $m\in N^{+}$. Then, for every environment e for each $\mathcal{S}\in\mathbf{K}$, there is a set $\{\Psi,\Phi^1,\ldots,\Phi^m\}$ of agents that satisfies Definition 9. It follows immediately that $\mathbf{K}\in\mathbf{Co}[\Psi]^{m}\mathbf{Ex}^{\pi}$. $\dashv$

The sketched proof of Proposition 2 highlights the meaning of the parameter m in the paradigm. For $\mathcal{S}\in\mathbf{Co}[\Psi]^{m}\mathbf{Ex}^{\pi}$, m is the number of the environments for $\mathcal{S}$ that Ψ does not $\mathbf{Ex}^{\pi}$-solve. Intuitively, Proposition 2 says that every agent that partially solves a structure is potentially able to fully solve it by looking for cooperation. The process of finding an helpful set of agents starts with a requirement of consistency. Then, any unsuccessful agent has only to recognize that communication and cooperation is better than isolation. As a corollary, it can be shown that for particularly "difficult problems" there is no team of helpful *computable* agents.

4 CONCLUDING REMARKS

We have demonstrated that nothing is gain in cooperative problem solving w.r.t. the class of solvable problems according to a fixed paradigm of solvability. Nevertheless, inductive cooperation has been proved to serve as an useful paradigm for cooperative actions to understand how unsuccessful agents can improve their problem-solving ability by jointly solving a problem. The paradigm of team-solvability we proposed is a paradigm of teamwork in a strict sense, where cooperation is formalized accordingly.

As far as we know, this is the first attempt to introduce cooperation within the framework of the model-theoretic tradition of *Formal Learning Theory* [16, 9]. A proposal has been recently put forth in AI that is quite related to ours in spirit, though not in the formal development. That is [29]. The use of cooperation in problem solving appears in [29] similar to what presented here. Some differences have been pointed out directly in the text.

A number of important directions remain to be pursued. The extension of $\mathbf{Co}^{\pi}$-solvability to teams of m agents for $m>2$ is one. It is also important to develop a paradigm of teamwork for "rational" agents, a paradigm that would explain how agents' beliefs about a given problem in a given environment (which includes the behavior of others insofar as it affects each agent's decisions) evolve until the agents have come to agree with the actual solution of the problem.

[6] Though agents' preferences may be integrated in the model as a *preference relation* (preorder) on agents' background world.

[7] See for instance [5] for the notion of elementarily equivalent structures.

The supplementation of rational choice theory we require is a theory of belief formation in social processes like cooperation, that is, a theory of rational cooperative problem solving.

ACKNOWLEDGMENTS

I thank Dick de Jongh and Franco Montagna for useful feedback on cooperation, and Fausto Giunchiglia for some quite meta-theoretical but winning advice.

REFERENCES

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

[2] A. Agostini, D. de Jongh, and F. Montagna. Coordination of 01 agents vs. coordination of worlds-based agents. To appear in ILLC prepublication series, Amsterdam, 2000.

[3] L. Blum and M. Blum. Toward a mathematical theory of inductive inference. *Information and Control*, 28(2):125–155, 1975.

[4] C. Boutilier. Planning, learning and coordination in multiagent decision processes. In *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge (TARK-96)*, pages 195–210, Amsterdam, 1996.

[5] C.C. Chang and J.M. Keisler. *Model Theory - 3rd edition.* North Holland, 1990.

[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] E. Martin and D. Osherson. Scientific discovery based on belief revision. *Journal of Symbolic Logic*, 62(4):1352–1370, 1997.

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

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

[14] S. Morris and H. S. Shin. Approximate common knowledge and coordination: Recent lessons from game theory. *Journal of Logic, Language, and Information*, 6:171–190, 1997.

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

[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. A universal inductive inference machine. *Journal of Symbolic Logic*, 56(2):661–672, 1991.

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

[19] D. Osherson and S. Weinstein. On the danger of half-truths. *Journal of Philosophical Logic*, 24:85–115, 1995.

[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 (AAAI-94)*, 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] M. P. Singh. The intentions of teams: Team structure, endodeixis, and exodeixsis. In *Proceedings of the Thirteenth European Conference on Artificial Intelligence (ECAI-98)*, pages 303–307, Brighton, 1998.

[25] R. J. Solomonoff. A formal theory of inductive inference. *Information and Control*, 7:7–22, 1964.

[26] G. Weiß. Learning to coordinate actions in multi-agent systems. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence (IJCAI-93)*, pages 311–316, Chambery, France, 1993.

[27] 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.

[28] M. J. Wooldridge and N. R. Jennings. Formalizing the cooperative problem solving process. In *Proceedings of the Thirteenth International Workshop on Distributed Artificial Intelligence*, pages 403–417, 1994.

[29] M. J. Wooldridge and N. R. Jennings. The cooperative problem-solving process. *Journal of Logic and Computation*, 9(4):563–592, 1999.