ECAI 2010
H. Coelho et al. (Eds.)
IOS Press, 2010


doi:10.3233/978-1-60750-606-5-163

# Learning and Meta-Learning for Coordination of Autonomous Unmanned Vehicles

## A Preliminary Analysis

**Predrag T. Tošić and Ricardo Vilalta**[1]

**Abstract.** We study models of coordination, negotiation and collaboration in *multi-agent systems* (MAS). More specifically, we investigate scalable models and protocols for various distributed consensus coordination problems in large-scale MAS. Examples of such problems include conflict avoidance, leader election and coalition formation. We are particularly interested in application domains where robotic or unmanned vehicle agents interact with each other in real-time, as they try to jointly complete various tasks in complex dynamic environments, and where decisions often need to be made "on the fly". Such MAS applications, we argue, necessitate a *multi-tiered* approach to learning how to coordinate effectively. One such collaborative MAS application domain is ensembles of *autonomous micro unmanned aerial vehicles* (micro-UAVs). A large ensemble of micro-UAVs on a complex, multi-stage mission comprised of many diverse tasks with varying time and other resource requirements provides an excellent framework for studying multi-tiered learning how to better coordinate. A variety of tasks and their resource demands, complexity and unpredictability of the overall environment, types of coordination problems that the UAVs may encounter in the course of their mission, multiple time scales at which the overall system can use learning and adaptation in order to perform better in the future, and multiple logical and organizational levels at which large ensembles of micro-UAVs can be analyzed and optimized, all suggest the need for a multi-tiered approach to learning. We outline our theoretical and conceptual framework that integrates reinforcement learning and meta-learning, and discuss potential benefits that our framework could provide for enabling autonomous micro-UAVs (and other types of autonomous vehicles) to coordinate more effectively.

## 1 INTRODUCTION AND MOTIVATION

*Coordination* is among the most important problems in *Distributed AI* research that studies various models and applications of multi-agent systems. According to [26], *multi-agent coordination* can be defined as "managing inter-dependencies among the activities of different agents". These interdependencies of agent activities can be of various types [26] (pp. 200 – 202), and the type of interdependencies is one of the factors determining the appropriate multi-agent coordination paradigm. Moreover, various interdependencies and hence the need to coordinate may arise both among self-interested, competitive agents, and the cooperative, distributed problem solving agents that typically share the same goals or objectives and are not competing with each other.

There are several common types of coordination problems that have been extensively studied in the MAS literature. Among them, various *distributed consensus* problems are particularly prominent. An important distributed consensus problem is *coalition formation* (e.g., [1, 4, 5, 9-13, 18-20]). In this paper, for space constraints reasons, we focus on a particular coordination distributed consensus problem – that of coalition formation among multiple collaborative autonomous agents that are *cooperating* with each other in order to better perform on their tasks [11-13, 18-21]. While the problem setting we consider is simpler than the one where the autonomous agents are *self-interested* and *competitive* (and hence, in general, may engage in both competition and cooperation with each other), as we shall see this setting still provides an abundance of research challenges – especially when it comes to addressing the appropriate ways of enabling these cooperative agents to *learn* how to coordinate more effectively based on their past interactions.

In many important *collaborative* MAS applications, autonomous agents need to dynamically form groups or coalitions in an efficient, reliable, fault-tolerant and partly or entirely distributed manner. In most of the literature on distributed coalition formation, agents' negotiation and coalition formation strategies are *static*, in that the agents do not learn from past experience, nor adapt their strategies to become more effective in their future coalition formation interactions. Moreover, most of the prior research that does consider learning and adaptation to improve coalition formation cf. focuses on various models of *reinforcement learning* (RL) at the level of *individual agents.*

We have recently proposed an integrated, multi-tiered approach to multi-agent learning on how to coordinate that systematically addresses agents' adaptability and ability to improve at forming coalitions in typical complex (in particular, noisy, partially observable and dynamic) environments. This integrated approach combines *reinforcement learning* (RL) of individual agents with *co-learning* and *meta-learning* at higher organizational levels [22]. In particular, we argue that the interaction and synergy between reinforcement learning of individual agents and meta-learning at the system level would enable agents to considerably improve in their coordination capabilities. The potential benefits of multi-tiered approach to learning in general, and of application of meta-learning techniques in particular, are especially significant when there is only a modest amount of past experience, limited computational resources of each individual agent (hence imposing constraints on how much can an agent spend on RL), and/or when there are

1 Department of Computer Science, University of Houston, Houston, Texas, USA. Email: *ptosic@uh.edu, vilalta@cs.uh.edu*

considerable changes in the agents' dynamic environments in which future coordination interactions and coalition formation attempts will take place.

The rest of the paper is organized as follows. First, we discuss the general problem of multi-agent coordination, and the more specific problem of coalition formation, in collaborative MAS; our focus is on motivating the need for various learning techniques in order to improve coordination and coalition formation abilities of agents over time. We then briefly discuss two important learning paradigms applicable in this context, those of *reinforcement learning* (RL) and *meta-learning* (ML). We find that, while RL has a relatively considerable prior art in the existing MAS literature, ML in the context of multi-agent learning is largely an unchartered territory. We briefly compare and contrast reinforcement learning at the level of individual agents and meta-learning at the large agent ensemble or system level. We then turn to an integrated, multi-tiered approach to multi-agent learning on how to coordinate and form coalitions more effectively, and how it can be fruitfully applied to an application domain that one of us has extensively studied in the past, namely, the large-scale ensembles of *micro unmanned aerial vehicles* (micro-UAVs) on multi-task missions carried out in highly dynamic, unpredictable, nondeterministic and partially observable environments [17, 18, 21].

## 2 AGENT COORDINATION AND COALITON FORMATION IN COLLABORATIVE MAS

As outlined in the introduction, distributed coordination and coalition formation in collaborative multi-agent domains are important problems that have been extensively studied by the *Distributed AI* research community. There are many important collaborative MAS applications where autonomous agents need to form groups, teams or coalitions. These applications include both purely software agents as well as robotic agents; the latter include, but are not limited to, various types of autonomous unmanned vehicles [17-21, 23, 27]. Collaborative agents may need to form teams or coalitions in order to share resources, complete tasks that exceed the abilities of individual agents, and/or improve some system-wide performance metric such as the speed of task completion [8, 12, 21].

One well-studied general problem domain is a collaborative MAS environment populated with distinct tasks, where each task requires a tuple of resources on the agents' part in order for the agents to be able to complete that task [11, 12, 18-21]. In this distributed task allocation context, agents need to form coalitions such that each coalition has sufficient cumulative resources or capabilities across the coalition members in order to be able to complete the assigned task.

There are at least two fundamental properties shared by many practical MAS applications that require the agents to be adaptable and capable of learning how to improve their coalition formation strategies. First, the same larger group of agents may need to engage in coalition formation or other types of coordination interactions with each other repeatedly, and for the purpose of effectively coordinating in the same or similar kind of environment, they need to become effective at completing the same or similar set of repetitive tasks. Clearly, being able to learn from past experience and then improve in future coordination interactions would be very beneficial.

Second, most realistic MAS environments, including those where coalition formation naturally arises as a way of solving the distributed task or resource assignment problem as outlined above, are characterized by a number of possible sources of uncertainty and noise [3, 4, 8, 21, 22]. Sources of uncertainty and noise may include (i) inaccuracies and inconsistencies in different agents' estimates of the tasks' utility values and/or resource requirements, (ii) a possibility of an agent's failure while working on a task as a part of one's current coalition, and (iii) inaccurate, incomplete and/or inconsistent estimates of individual agents' abilities and their potential contribution as members of various coalitions, that is, *imperfect beliefs* about other agents [4, 17, 18]. Once these sources of uncertainty are taken into account, and assuming agents would need to form coalitions repeatedly, clearly each agent should be able to learn how to better identify which candidate coalitions have a high(er) chance of success, i.e., are most likely to succeed at completing future tasks.

We argue that a need for learning arises naturally in this kind of noisy, imperfect information collaborative MAS environments with repeated coalition formations and coalition-to-task assignments at two qualitatively distinct levels. At one level, we find learning to identify individual agents that among their peers are better (more reliable and effective) coalition partners than others. In most scenarios we have considered or found in the existing literature, this individual agent learning is of the *reinforcement* nature: an agent learns based on the past track record of rewards from various completed tasks, which were accomplished while the agent was a member of various coalitions. At a different level, we encounter a kind of learning that takes place at the 'system level' or agent ensemble level. For example, how would the MAS designer (e.g., in team robotics applications) or the central command-and-control (e.g., in emergency response, military or law enforcement applications) go about re-defining or modifying the agents' coalition formation strategies, the incentives given to the agents to form various coalitions, and how to reconcile inconsistencies of different agents' views of the world, in order to make the future autonomous coalition formation process among its agents as effective as possible?

We make the case that what is really needed is to combine reinforcement learning models and techniques with those of meta-learning. In our view, only a multi-tiered approach to learning and adaptation in multi-agent coordination in general, and distributed coalition formation in particular, holds a true promise for making a breakthrough on the fundamental challenge of collaborative MAS research, that of *learning how to coordinate effectively*. In particular, we posit that the problem of learning how to coordinate fundamentally needs to be tackled both at the level of individual agents and at the overall MAS level [22].

## 3 REINFORCEMENT LEARNING IN AGENT COORDINATION & COALITON FORMATION

During any coordination encounter in general, and the coalition formation process in particular, an agent encounters various sources of uncertainty and noise that affect its effectiveness as well as its preferences over possible candidate coalitions with other agents. Uncertainty and noise may affect the following:

(i) Agent's perception of various tasks, and in particular tasks' (i) utility values (to the agent and/or to the entire system) and (ii) resource requirements (i.e., how difficult is it going to be to complete those tasks);

(ii) Agent's perception of other agents, in particular those other agents' capabilities and reliability as coalition members;

(iii) Inconsistencies in task preferences (e.g., in terms of different utility evaluations of a given task by different agents, or different estimates of that task's resource requirements) by different members of a potential coalition of agents.

In many applications, the same ensemble of agents may need to perform multiple stages of coalition formation and coalition-to-task mapping. Each member of such agent ensemble, therefore, could benefit from being able to learn which coalition partners are more reliable or useful than others, based on past experience. In most situations, learning is of the *reinforcement* nature: rather than being provided clues by an outside teacher, an agent receives, in general, different payoffs for different choices of coalition partners and of tasks that the formed coalitions are mapped to. These differences in payoff outcomes are the result of varying effectiveness of different coalitions that this agent forms at different stages of the MAS deployment [4, 5, 22].

Several *reinforcement learning* (RL) models in the context of multi-agent coalition formation have been studied. We briefly discuss some representative research directions found in the existing literature that are directly relevant to our problem setting, namely to locally-bounded robotic and unmanned vehicle agents that are *collaborative* and hence strictly *cooperative*, never *competitive*, with respect to each other. We observe that self-interested agents that are, in general, competitive with each other, may still need to engage in various forms of coordination and even some (typically limited) forms of cooperation [25]. More details on various collaboration and cooperation challenges in multi-agent systems can be found in [25] and [26].

While we assume that each agent has a local picture of the world, and therefore different agents' pictures of the world are in general going to differ from each other, we study agents that do not have conflicting interests, nor selfish agendas, and hence no interest in trying to out-smart or "out-play" each other.

In particular, the rich prior art that addresses reinforcement learning (or other types of learning) in multi-agent domains where the agents are assumed *self-interested*, and where agents therefore in general *compete* (or possibly both cooperate and compete) with each other, is not directly relevant to us. An excellent survey on rational distributed decision making among *self-interested agents*, including various aspects of negotiation and coordination, can be found in Chapter 5 of [25]. However, we are only interested in those aspects of coordination and coalition formation that are of relevance to the agents that, while locally constrained, are *not self-interested*.

An important critical survey of the broad general area of multi-agent reinforcement learning, that in particular identifies four "core" categories of problems in multi-agent learning, is that of Shoham et al. [14]. Our problem, which is collaborative multi-agent learning on how to coordinate and form coalitions effectively, would fit into the second category in the MAS RL problem taxonomy proposed in [14].

While our focus on *strictly collaborative* MAS considerably narrows down the overall problem setting (since we explicitly exclude the competitive MAS from the onset), the collaborative multi-agent domains still offer a wealth of interesting problems. The brief discussion of prior art on RL for multi-agent coordination, coalition formation and task allocation in collaborative MAS that follows is not meant to be exhaustive, but rather illustrative of some interesting approaches to reinforcement learning in the above-mentioned contexts.

In [4], Bayesian models of RL for coalition formation in the presence of noise are proposed. Each agent maintains its explicit beliefs about properties of other agents, that is, agents engage in *mutual modeling* [26]. These beliefs are then refined and updated based on an agent's experience, i.e., on prior outcomes resulting from repeated multi-agent interactions. Each agent is learning to control a stochastic environment which is modeled as a *Markov Decision Process* (MDP). Research direction in [1] assumes an underlying organizational structure of the multi-agent system, and a distributed coalition formation process that is guided by that structure. The proposed approach uses RL techniques to improve upon local agent decisions within the larger organizational context; however, the learning still takes place at the level of individual agents. In [5], agents use *case-based learning and reasoning* in order to make and maintain simple models of each other, and then engage in reinforcement learning at both *joint-behavior* and *individual-behavior* levels. The joint-behavior learning proposed in [5] is an interesting and promising approach that has a distinctly *co-learning* flavor; the main challenge there is scaling up such joint behavior learning from a handful of agents to dozens to hundreds or more agents. We discuss co-learning mechanisms and their strengths and limitations in some detail in the next two sections.

Several other approaches based on RL at the level of an *individual agent* in order to improve effectiveness of coalition formation have been studied. More detailed surveys of the state-of-the-art of RL in the context of coalition formation can be found, e.g., in [5] and [8].

# 4 META-LEARNING FOR COORDINATION AND COALITION FORMATION

Various forms of reinforcement learning, as briefly discussed in the previous section, pertain to how an *individual agent* can adapt and improve its coalition formation strategy and selection of coalition partners. However, learning from past experience can take place at higher organizational levels than that of individual agents; in particular, it can take place at the system level, as well.

Depending on the nature of MAS, this system level learning could refer to, e.g., self-organizing adaptability of agent ensembles or to meta-learning of the MAS system designer or other central authority. For example, in case of a collaborative MAS application of a system of autonomous *micro unmanned aerial vehicles* (micro-UAVs) on a complex, multi-stage, multi-task mission [17, 18], this higher-level learning could take place at the central command-and-control.

In contrast to the relatively rich literature on individual agent's RL in the context of various MAS coordination, prior art on *meta-learning* [2, 24]) applied to improving coordination among collaborative agents is very modest. Reference [16] studies meta-learning processes in MAS among self-interested agents that are *competing* with each other, as opposed to engaging in cooperative distributed problem-solving. This work focuses on algorithmic game-theoretic aspects of multi-agent interactions. In that context, a number of assumptions are made that are not suitable for our problem setting, including (i) competitive nature of inter-agent interactions and (ii) small, *a priori* known (and fixed) finite sets of available actions to each agent at each "move" of the "game". Furthermore, what [16] refers to as *meta-learning* is more properly described as agent *co-learning*.

The closest in spirit to our approach to learning and meta-learning for more effective coordination and coalition formation in *collaborative* distributed problem solving MAS is found in [15], which addresses learning how to improve coalition formation at different organizational levels for general MAS that need not be strictly collaborative. [15] studies learning at what the authors refer to as *tactical* and *strategic* levels. At the tactical level (of an individual agent's decision-making), *RL* is used to identify most viable candidates for coalition partners, whereas *case-based learning* (CBR) is used to refine specific negotiation strategies used by an individual agent. At the strategic level, a distributed, cooperative CBR aids in improving the overall negotiation capabilities, thereby leading to a more effective coalition formation. However, [15] doesn't consider meta-learning techniques at all.

We outline how meta-learning could enable a better multi-agent coordination and more effective, adaptable and efficient coalition formation at the system or *strategic* level [15, 22]. Past performance and coordination strategies (such as choice of coalition partners, tie-breaking mechanisms, and how successful various resulting coalitions were in performing their tasks) of collaborative MAS can be captured in a meta-dataset, that would be stored at a (typically, central) *knowledge base* (KB). Such meta-dataset would contain various parameters that are used by the agents during the coalition formation process, where selected values of these parameters, in general, are associated with different levels of coalition formation efficiency and/or subsequent coalition successfulness. A meta-learning system can exploit this meta-knowledge to learn to associate various parameters with successfulness. This meta-learning system would use the large system-level KB with complex data sets in order to make complex, typically probabilistic/statistical inferences about the future effectiveness of various coalition formation strategies and choices, based on the past histories (i.e., cumulative experience of all individual agents in the system). Such cumulative experience and inferences based on that experience can be then exploited to adjust how agents select future coalition partners, as well as to dynamically adapt coalitions and their overall capabilities to tasks and their resource requirements. In the fairly common MAS scenario where an agent ensemble repeatedly engages in coalition formation and coalition-to-task mapping interactions over a considerable time, the accumulated experience can reveal statistically relevant patterns to suggest the best coalition formation strategy for particular tasks [24].

In most practical scenarios of our interest (such as team robotics, ensembles of micro-UAVs, as well as various *swarm intelligence* applications), accumulating and storing all this experience across large agent ensembles, as well as making non-trivial statistical inferences based on the knowledge base created from that stored experience, would likely be beyond the computational resources of any single agent [21, 22]. Moreover, both creation and subsequent use of such a system- or ensemble-level KB would also likely exceed the joint resources or abilities of smaller groups of agents, as well; as such, inferences at this level, and presumably the improved coordination abilities based on those inferences, would therefore also be beyond what is achievable via co-learning at the level of small groups of agents. Therefore, complex inferences at the level of large agent ensembles or the entire system should not be expected to be feasible to achieve at lower organizational levels of MAS, nor via the classical reinforcement learning and/or co-learning mechanisms *alone.*

We further argue that, in many robotic and autonomous unmanned vehicle applications, meta-learning is indeed necessary if the system designer hopes to take maximal advantage of the historical records of her MAS system performance. Furthermore, resources necessary for successfully undertaking such a meta-learning approach are usually readily available – at least insofar as *offline* learning and inference are concerned. The results of such offline learning can then be made available to the agents as those agents repeatedly engage in the same, or similar, type(s) of interactions. In particular, in most practical MAS applications that we have studied, complex statistical pattern inference would certainly exceed the available computational resources of individual agents; hence, the potential benefits of such inference capabilities wouldn't be expected to be attainable without meta-learning at the system level [22].

We conclude our discussion of the role meta-learning by summarizing the main conceptual, logical and architectural differences between reinforcement learning and meta-learning in collaborative, distributed problem-solving MAS.

*Reinforcement learning* is typically done: (i) at the level of a single agent; (ii) within the agent (and its memory, processing, sensing, etc. resources); (iii) *online* (and, in many applications, esp. those of team robotics and autonomous unmanned vehicle nature, in real-time); (iv) is resource-bounded; and (v) is based on local information and knowledge available to a single agent.

In contrast, *meta-learning* in our proposed overall learning architecture would be done (i) at the system level (see above for some examples, what that could mean in practice); (ii) logically as well as architecturally *externally* to individual agents; (iii) *offline*; (iv) with an access to much richer data sets, as well as to more processing power and other computational resources; and (v) would not be subject to local constraints of individual agents.

In particular, the "meta-knowledge" that is used in the course of meta-learning would be stored in a (potentially, very sizable) knowledge base that is *external* to the agents themselves. The ability to create and maintain an offline KB external to the agents, and to perform complex meta-learning and meta-reasoning analysis with that KB by centralized resources available to the MAS designer but not to her individual agents, may well be the most practical (or even the only possible) way of taking advantage of the rich, complex, large data sets that capture detailed histories of the past agent interactions. With such a large knowledge base, and a sufficiently powerful inference engine, it would be possible to represent and analyze knowledge and meta-knowledge about (i) all individual agents and their past actions, deliberations and performances, (ii) local as well as non-local inter-agent interactions, and (iii) global properties of the large-scale agent ensembles and their environments.

## 5 MULTI-TIERED LEARNING OF HOW TO BETTER COORDINATE MICRO-UAVs

A collection of *micro unmanned aerial vehicles* (micro-UAVs) that are autonomous (in particular, not remotely controlled by either a human operator or a computer program) and that need to coordinate with each other in order to accomplish a complex, multi-task mission in a highly dynamic, unpredictable, partially observable environment provide an ideal tested for modeling, designing and analyzing large-scale collaborative MAS operating in "the real world". Such ensembles of micro-UAVs can be used for various surveillance, reconnaissance, search-and-rescue and other similar tasks, including longer-term missions made of a variety of such tasks [17, 18]. Micro-UAVs are, in general, equipped with sophisticated sensors (radars, infra-red cameras etc.), actuators or "payloads", and communication links (typically, radios). Their communication may include peer-to-peer message exchanges, local broadcasts or multicasts, global broadcasts/multicasts, and message exchanges with centralized command-and-control [6, 7, 17, 18].

The coordination problems encountered by an ensemble of micro-UAVs can range from conceptually simple (but challenging in practice) collision avoidance to distributed divide-and-conquer "single shot" task allocation to complex fully or partially distributed planning [19]. Some consensus problems that naturally arise in UAV deployments that are fully distributed (i.e., when no communication with command-and-control is possible or feasible) include coalition formation and leader election [19, 21]. We point out that, while coordinating unmanned vehicles originated as primarily a military application, coordination of dozens or even hundreds of autonomous unmanned vehicles of various types is a problem of increasing importance in the private sector and various industries, as well. An interested example on unmanned vehicles used in very large warehouses, and DAI research behind enabling those vehicles to work together without colliding with each other, can be found in [27].

A large micro-UAV ensemble on a complex, multi-stage mission comprised of diverse tasks with varying time and other resource requirements provide an excellent context for multi-tiered learning on how to better coordinate [17, 22]. A variety of tasks and their resource demands, complexity of the overall environment, a variety of coordination problems that the UAVs may encounter in the course of their mission, multiple time scales at which the overall system can use learning and adaptation in order to perform better in the future, and multiple logical and organizational levels at which large such micro-UAV ensembles can be analyzed and optimized, all suggest the need for a multi-tiered approach to learning. At the level of an individual UAV, the familiar reinforcement learning paradigm is suitable. Due to space constraints, we won't discuss it further; we will focus, instead, on co-learning and meta-learning in the outlined setting.

*Co-learning* among small groups of UAVs could take place along similar lines to what is proposed in [15]. One caveat is that the need for one agent to model some of the other agents explicitly would not be motivated by differing, possibly conflicting, interests of different agents. Instead, it would be due to any combination of the following: (i) imperfections of communication links, (ii) inaccuracies in how agents evaluate tasks and, in particular, the suitability of their own capabilities or resources to perform those tasks, (iii) inconsistencies in perceived value and resource requirements of a task as seen by different agents, (iv) different capabilities of agents, and (v) agents' inconsistent beliefs about each other's capabilities.

Consider a simple example: an agent, A, identifies some task *T* that A estimates would require two agents of A's capabilities to complete. Among near-by UAVs, A can pick UAV B or UAV C to form a two-member coalition that would be assigned to task *T*. Ability of A (and other agents, including B and C) to *co-learn* would enable agent A to (i) solicit feedback from B and C on how they view task *T*'s value and resource requirements, and to compare those with its own view of the task, (ii) based on past interactions with B and C, to have a preference for one over the other as a coalition partner, (iii) to learn from B and C if they happen to have identified other tasks worth completing, (iv) to have a degree of trust or confidence in B's and C's evaluations of their own abilities, as well as (v) of the values and resource requirements of other tasks that agents B and C may be interested in. Based on (i) – (v), agent A may be able to make a more informed decision on matters such as (a) whether to still pursue task T or opt for some T' that it learns about from B or C, (b) which of the alternative tasks (if more than one such T' exists) to choose, and (c) which coalition partner, B or C, to choose as preferred coalition partner for the task of choice.

Co-learning as above, however, could hardly be expected to scale up; that is, a micro-UAV can perhaps maintain explicit models of a handful of other micro-UAVs, but in case of a very large ensemble (made of hundreds or possibly thousands of such micro-UAVs), trying to model most or all of other agents would simply exceed the memory and processing power of an individual agent. Moreover, such swarm micro-UAV deployments would likely entail each UAV being able to directly interact with only a handful of others; moreover, flooding this *ad hoc* network of micro-UAVs with the global information (say, sent from the central command-and-control) would likely not work well either, both from the communication cost standpoint and from the computational processing stand-point.

In our view, to take the full advantage of accumulated global knowledge and meta-knowledge about all the agents in the system, their past interactions, various tasks and their properties, and successfulness of different previously used coordination strategies, a genuine *meta-learning* [2, 24] approach is required. Due to its memory and CPU time resource requirements, this meta-learning would be expected to take place offline, perhaps at a centralized command-and-control. We briefly outline (i) the potential benefits of integrating such offline, computationally intensive meta-learning (and meta-reasoning) with the online, real-time (RT) reinforcement learning and co-learning that are less computationally demanding but also provide less insights, and (ii) what kind of knowledge would be stored in an appropriate knowledge base that a meta-learning and meta-reasoning engine would then use in order to enhance longer-term learning and decision-making.

Consider a scenario where the command-and-control may want to learn from the experience with the past deployments of UAV teams, so that the future deployments of a similar nature don't repeat the same mistakes and hopefully achieve the required coordination faster than before. To do this, without meta-learning, the command-and-control would have to base its inferences on what went wrong, what needs to be improved (and how) entirely on human experience and expertise (i.e., the military or law enforcement commanders looking into videos, or log files, or other track records, and then deciding how to re-program and re-deploy the UAVs in the future). Meta-learning and meta-reasoning would enable partial (or, in principle, even complete) automation of such future deployments, where lessons are learned from the past mistakes at the level of the entire system, not just individual UAVs; hence, this level of automated learning and reasoning would go beyond what individual UAVs can reinforcement-learn from their local environments, histories etc.

Meta-learning and meta-reasoning could indicate what coordination mechanisms have better chance of success for a given terrain or type of mission or types/capabilities of given UAVs than other; but the future deployments of UAVs would still be genuinely distributed (i.e., no remote control), except that some new knowledge and meta-knowledge is periodically "built into" the agents; and indeed some of that new knowledge may have been obtained in a centralized manner, as discussed above.

The kinds of meta-data that the *knowledge base* (KB) at the command-and-control center would store could include properties of all tasks encountered by any of the micro-UAVs so far, individual experiences of agents across the entire large ensemble and across different time epochs, and various meta-knowledge inferred from the historical records. Some examples of such records would be those pertaining to appropriate agent coalitions, resources that were required and how much time it took in the past to complete particular type of tasks (such as search-and-rescue or surveillance or reconnaissance tasks) in particular types of environment (for example, mountainous terrain vs. flat terrain, the size in square miles of the land area across which the search took place, what were the weather conditions, was resistance from an adversary encountered and what type of resistance, etc.). Such meta-knowledge could then be used in the second stage of the meta-learning inference engine's operation to provide the agents with summaries of all the past experience in forms of task and candidate coalition rankings, "bonus" incentives to build coalition with one subset of agents instead of another, revised estimates of the values (expected payoffs) and resource requirements of new tasks that are sufficiently similar to some of the previously encountered tasks, and so on.

# 6 SUMMARY

We survey in this paper distributed coordination and coalition formation in collaborative MAS, and discuss the need for multi-agent learning in order to improve coordination. We propose a novel multi-tiered approach to multi-agent learning, where learning from past interactions takes place at different logical and organizational levels, from individual agents to the entire large-scale agent ensembles and the system designer level. In particular, we argue that meta-learning applied to coordination and coalition formation, and integrated with individual agents' reinforcement learning and co-learning, holds a great promise for genuinely adaptable, distributed problem solving multi-agent systems whose complex behaviors need not be (and often times in practice, actually *cannot* be) "hard-wired" at the design time.

We then focus on an application, namely, micro-UAVs. We study multi-tiered learning in that setting, and outline how would meta-learning apply to UAV coordination. Inferring useful meta-knowledge and then using that meta-knowledge to help agents revise their beliefs and intentions, and ultimately to coordinate better, would in most situations be beyond the resources of individual agents such as micro-UAVs. Therefore, large-scale micro-UAV deployments, as well as many collaborative large-scale MAS conceptually similar to micro-UAVs, could uniquely benefit from meta-learning and meta-reasoning techniques.

## REFERENCES

[1] S. Abdallah, V. Lesser. *Organization-Based Cooperative Coalition Formation, in* Proc. IEEE / WIC / ACM Int'l Conf. on Intelligent Agent Technology, IAT'04 (2004)

[2] P. Brazdil, C. Giraud-Carrier, C. Soares, R. Vilalta. *Metalearning: Applications to Data Mining,* Springer (2009)

[3] Chalkiadakis, G., Boutilier, C. *Coordination in Multiagent Reinforcement Learning: A Bayesian Approach*, in Proc. Autonomous Agents & Multi-Agent Systems AAMAS'03 (2003)

[4] Chalkiadakis, G., Boutilier, C. *Bayesian reinforcement learning for coalition formation under uncertainty*, in Proc. AAMAS'04, New York City, New York, USA (2004)

[5] X. Li, L.K. Soh. *Investigating reinforcement learning in multiagent coalition formation*, TR WS-04-06, AAAI Workshop Forming and Maintaining Coalitions & Teams in Adaptive MAS (2004)

[6] M-W. Jang, et al. *An Actor-based Simulation for Studying UAV Coordination*, Proc. 16th Euro. Simulation Symposium (ESS '03), pp. 593-601, Delft, The Netherlands (2003)

[7] M-W Jang, G. Agha. *On Efficient Communication and Service Agent Discovery in Multi-agent Systems*, Proc. 3rd Int'l Workshop on Software Engin. for Large-Scale MAS (SELMAS'04), Edinburgh, Scotland (2004)

[8] de Oliveira, D. *Towards Joint Learning in Multiagent Systems Through Opportunistic Coordination*, PhD Thesis Proposal, Univ. Federal Do Rio Grande Do Sul, Brazil (2007)

[9] T.W. Sandholm, V.R. Lesser. *Coalitions among computationally bounded agents*, in *Artificial Intelligence* vol. 94, pp. 99-137 (1997)

[10] T. W. Sandholm, K. Larson, M. Andersson, O. Shehory, F. Tohme. *Coalition structure generation with worst case guarantees*, in *AI Journal* vol.111 (1-2), pp. 210-238 (1999)

[11] O. Shehory, S. Kraus. *Task allocation via coalition formation among autonomous agents*, in Proc. IJCAI-95, Montréal, pp. 655–661 (1995)

[12] O. Shehory, K. Sycara, S. Jha, *Multi-agent coordination through coalition formation*, in *Intelligent Agents IV: Agent Theories, Architectures and Languages*, LNAI, vol. 1365, pp. Springer (1997)

[13] Shehory, O., Kraus, S. *Methods for task allocation via agent coalition formation,* in *AI Journal,* vol. 101 (1998)

[14] Shoham, Y., Powers, R., & Grenager, T. *Multi-Agent Reinforcement Learning: A Critical Survey*, Technical Report (2003)

[15] Soh L. K., Li, X. *An integrated multilevel learning approach to multiagent coalition formation*, in Proc. Int'l Joint Conf. on Artificial Intelligence IJCAI'03 (2003)

[16] R. Sun. *Meta-Learning Processes in Multi-Agent Systems*, in *Intelligent agent technology: research and development*, N. Zhong, J. Liu (eds.), pp. 210-219, World Scientific, Hong Kong (2001)

[17] P. Tosic, G. Agha. *Understanding and Modelling Agent Autonomy in Dynamic Multi-Agent, Multi-Task Environments*, in Proc.1st Euro. Conf. Multi-Agent Systems (EUMAS'03), Oxford, England (2003)

[18] P. Tosic. *Modeling a System of UAVs on a Mission*, invited session within 7th World Multiconference on Systemics, Cybernetics & Informatics (SCI'03), Orlando, Florida (2003)

[19] P. Tosic, G. Agha. *Maximal Clique Based Distributed Coalition Formation for Task Allocation in Large-Scale Multi-agent Systems,* Post-Proc. MMAS'04, Springer LNAI series, vol. 3446 (2005)

[20] P. Tosic, G. Agha. *Maximal Clique Based Distributed Group Formation Algorithm for Autonomous Agent Coalitions*, Workshop on Coalitions and Teams (W10), in AAMAS'04, New York City, NY (2004)

[21] P. Tosic. *Distributed Coalition Formation for Collaborative Multi-Agent Systems,* MS thesis, Univ. of Illinois at Urbana-Champaign (UIUC), Urbana, Illinois, USA (2006)

[22] P. Tosic, R. Vilalta. *A unified framework for reinforcement learning, co-learning and meta-learning how to coordinate in collaborative multi-agent systems,* Proc. Int'l Symp. Computer Science (ICCS-2010), Amsterdam, Procedia Comp. Sci, Elsevier (2010)

[23] L. Vig, J. Adams. *Issues in multi-robot coalition formation*, in Proc. Multi-Robot Systems: From Swarms to Intelligent Automata, vol.3 (2005)

[24] Vilalta, R., Giraud-Carrier, C., Brazdil, P. *Meta Learning Concepts and Techniques* (2005)

[25] G. Weiss (ed.) *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence,* MIT Press, 1999

[26] M. Wooldridge. *An Introduction to MultiAgent Systems,* John Wiley & Sons (2002)

[27] P. R. Wurman, R. D'Andrea, M. Mountz. *Coordinating Hundreds of Cooperative, Autonomous Vehicles in Warehouses,* in *AI Magazine,* vol. 29 (1), Spring (2008)