International Workshop on Massively Multi-Agent Systems (MMAS2018)
in conjunction with IJCAI/ECAI/AAMAS/ICML 2018

Distributed Speaking Objects: a Case for Massive Multiagent Systems

Marco Lippi, Marco Mamei, Stefano Mariani, and Franco Zambonelli

Dipartimento di Scienze e Metodi dell'Ingegneria
Università di Modena e Reggio Emilia, Italy

Abstract. Smart sensors and actuators, embedding learning and reasoning features and associated to everyday objects and locations, will soon densely populate our everyday environments. Being capable of understanding, reasoning, and reporting, about what is happening (for sensors) and about what they can make possibly happen (for actuators), these "speaking objects" will thus be assimilable to autonomous situated agents. Accordingly, populations of speaking objects will define dense and massive multiagent systems, devoted to monitor and control our environments, let them be homes, industries or, in the large-scale, whole cities. In this context, the necessary coordination among speaking objects will be likely to become associated with the capability of argumenting about situations and about the current state of the affairs, triggering and directing proper distributed conversations, and eventually collectively reach future desirable state of the affairs. In this article, we detail the speaking object scenario, overview the key enabling technologies, and analyze the key challenges for engineering large-scale collectives of speaking objects and their conversations.

1 Introduction

The *Internet of Things* (IoT) is enabled by the possibility of enriching physical objects and places with wirelessly accessible sensing, computing, and actuating capabilities [3], such that everything in our physical and social worlds will become a node in a large-scale situated network, supporting coordinated actions to sense and control the world itself and to facilitate interactions with it [6].

As of today, most of the approaches to engineer IoT systems still consider IoT devices as simple providers of services, either sensing services producing *raw data* or actuating services executing specific *commands* [3]. From the architectural viewpoint, most approaches adopt a *centralized* cloud perspective: raw sensor data is collected at some control point, there analyzed to infer situations and events in the concerns of interest, and commands for the actuators are generated to have them produce some effect on the smart objects in the environment in which they situate. However, some recent technological evolutions [10, 1, 34] let us point to a novel scenario:

- IoT devices can and are going to become much smarter [10]. On the one hand, rather than simply producing streams of data, smart sensors can integrate Artificial Intelligence (AI) tools, thus becoming capable of *understanding* and reporting – via factual assertions and arguments – about what is happening around. On the other

hand, smart actuators will become increasingly autonomous and *goal-oriented*, and able to decide how to act towards the achievement of specific goals [1]. In other words, such smart objects are becoming de facto software agents or, as we like to call them, "speaking objects" [26].
– Multitudes of speaking objects will form the nodes of massive distributed multiagent systems that can be exploited to monitor and control activities in real-time in our everyday environment. Although centralized cloud-based approaches are here to stay for the sake of global data analysis and long-term planning, speaking objects will have to interact and coordinate with each other in a distributed way, to ensure prompt response to situations [34].

Clearly, the very nature of speaking objects will dramatically change the approaches to implementing and coordinating the activities of distributed processes. In fact, coordination is likely to become associated with the capability of *argumenting* about situations and about the current "state of the affairs" [10], by reaching a consensus on what is happening around and what is needed, and by triggering and directing proper decentralised semantic *conversations* to decide how to collectively act in order to reach future desirable state of the affairs.

Based on such envisioned scenario, the contributions of this paper are as follows:

– We analyze the key concepts behind speaking objects, and show how they are going to change the very nature of decentralized coordination, challenging traditional approaches to distributed computing and calling for novel *conversational* approaches.
– We overview the key technologies and approaches that, in such a novel scenario, will have to be involved in the engineering of systems and services, and will have to become core expertise for distributed systems engineering. Among the others, these include knowledge representation and *commonsense* reasoning, *machine learning*, *goal-oriented* programming, *argumentation* models and technologies, and *human-computer* interfaces.
– We identify some key research challenges that will have to be faced to pave the way towards a novel and effective approach for the engineering of these new classes of distributed systems. These include challenges at the level of software engineering models, middleware technologies, user involvement, control and understandability, security.

To ground the discussion with an exemplar case study, we will consider the case of a large-scale *smart hospital* instrumented to support health monitoring and assisted living [17]. We assume the hospital to be densely enriched with connected sensors and actuators, at the level of basic infrastructures (e.g., lightening, heating), all its rooms (with ambient cameras, controllable doors and windows), appliances (e.g., furniture, clocks, TV, fridge, etc.), and medical devices (e.g., spyrometers, Fitbits, etc.). This infrastructure, possibly including wearable bio and activity sensors, can be used to monitor the living and health conditions of patients, and to dynamically control the overall configuration of the hospital to fit peculiar needs and contingencies.

2 Speaking Objects as Cognitive Goal-oriented Agents

Currently, in the IoT arena (and in related typical application scenarios, from smart homes to smart cities and transportation) the concept of *smart object* is mostly associated to the possibility of attaching ICT devices to physical objects and places, thus turning them into: *(i) sensors*, capable of sensing a large amount of properties related to our physical/social worlds, and producing big streams of data to be collected at some centralized (or semi-centralized as in edge/fog computing approaches [40]) point for later analysis; *(ii)* remotely controllable *actuators*, capable of enacting specific configurations or actions in the surrounding environment, by receiving appropriate commands.

However, recent progress across many different areas indicates that smart objects are improving fast beyond such mere sensing and actuating capabilities, to become capable of cognitive goal-oriented behavior. That is, to become de facto autonomous agents.

2.1 From Data Sensing to Cognitive Sensing

The increasing computational power that can be embedded in sensors and everyday objects, along with advancements in machine learning techniques, is making it possible for smart objects to analyze *locally* the stream of sensed data in order to extract relevant features from it. A trivial example is a wrist band capable of associating the sensed patterns of movement of patients to actions like "walking", "running", "sleeping" (see Figure 1), or a control camera that detects the presence of specific objects in the recorded scene, such as "stretcher in corridor X". To some extent, such objects are already becoming "speaking", by evolving from producers of raw data streams (a capability that they nevertheless preserve) to producers of high-level concepts.

However, we can soon expect that such capabilities will evolve in order to recognize more complex situations, making objects capable of *causally* connecting individual patterns into composite situations, that is, making assertions about what is happening around them. For instance, a fitbit may construct the assertion that "Marco has

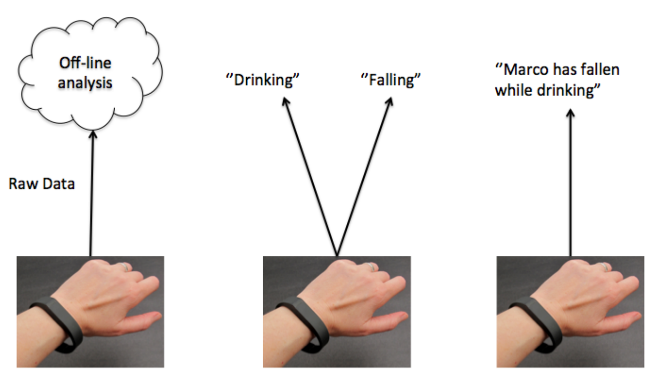

Fig. 1. From simple sensors to speaking objects.

fallen while drinking" from the sensing of two distinct patterns. Or a camera may perform scene understanding, by relating the individual objects it recognizes, e.g., "patient Marco has left the stretcher in corridor X". Such complex *situation recognition* is a hot topic for research in computer vision and in pervasive computing in general [39].

Further capabilities of asserting about complex situations arise from *sensor fusion* techniques, where the outputs of multiple sensors – each with a specific perspective on the surrounding world – are combined together to form a more comprehensive understanding. For example, fusing information from a camera and a temperature sensor in a smart room can eventually enable to assert that "the temperature is dropping down because the window is open".

Last but not least, the possibility for humans to enter the picture and act themselves as speaking objects (e.g., by posting information via their mobile phones), brings further possibilities of complex event recognition to the scenario.

We emphasize that our concept of *speaking objects* should not be interpreted solely as the capability of interacting via natural language (which nevertheless is an important feature in the overall framework, as we will discuss in the following) but more generally as the capability of expressing and understanding *assertions* about situations, regardless of the media and language via which they are delivered.

2.2 From Actuating Commands to Achieving Goals

Getting to actuators, our perspective is that smart actuating objects (capable of performing some action in the environment) will become capable of "hearing" what are the goals or situations to be achieved, and achieve them *autonomously*.

Again, we emphasize here that it is not a matter of having smart tools (such as Amazon Echo or Google Home) capable of interpreting vocal commands to activate some home appliances. In fact, whether triggered by vocal commands or by traditional service invocations, current appliances are simply interpreting *commands* and executing them. We are rather talking of moving from a command-based mode of operation to a *goal-based* one. Instead of telling actuators what to do, a goal-based approach relies on expressing a desirable *state of the affairs* to be achieved with respect to some environmental configuration, and let them autonomously evaluate what actions to make in order to reach it.

For instance, in the hospital scenario, a patient can simply express some desire (e.g., "I need to sleep") and have the light system start operating in autonomy, adjust lightning accordingly. However, autonomous goal-oriented smart objects will soon become pervasive. In a smart hospital scenario, besides goal-based robotic assistants, one can even think of simple smart goal-oriented appliances. For instance, a smart desk lamp that autonomously moves and tunes intensity to ensure optimal illumination in spite of changing environmental conditions [1].

Smart actuator objects, to achieve their goals, must acquire information about the current state of the affairs, which requires gathering information from smart sensors. Also, they must sometimes interact with each other and with non-smart objects (e.g., non goal-oriented actuators). For instance, in a smart hospital, in order to achieve specific temperature and humidity comfort levels, the A/C system might be in need to cooperate with the heating system and should be allowed to operate the opening/closing of the windows (assuming such windows as non goal-oriented).

The requirement of *interaction* brings us to the next section.

3 Distributed Coordination as a Conversation

In an environment populated by smart speaking objects (e.g., sensors) and by a variety of smart hearing objects (e.g., actuators), the issue of *coordinating* their distributed activities arises. In fact (see Figure 2):

- Speaking objects sense and have to produce an understanding of the situations around, for which they may be in need to exchange information (to complete information or disambiguate it).
- Speaking objects have to talk with hearing objects to inform them about what is happening (the current state of the affairs and the *reasons* causing them), which is necessary for hearing objects to plan actions.
- Hearing objects may have to talk to each other to agree on common courses of actions, whenever a desired state of the affairs (either embedded in their code or dynamically expressed at run-time) requires the cooperation of multiple actuators, or may be achieved in multiple ways by different actuators, or multiple conflicting views of the desired state of the affairs exist.
- All of which to form a *closed loop* [21], in which any action by the actuators produces some changes in the environment that have to be immediately sensed to provide feedback for the actuator themselves. Given such dynamics, and the possibility of expressing new desires in real-time, centralized (e.g., in the cloud) approaches become unsuitable, whereas *decentralized* coordination between the different objects (and possibly the concerned human actors) becomes mandatory, possibly with the support of some local hub [40].

In the following we show that, in the envisioned scenario, coordination between speaking and hearing objects naturally assumes the form of a distributed multi-party *conversation*, or *dialogue* [2], among autonomous agents.

3.1 From Coordination to Conversations

A *conversation* is a *session of interaction* between an ensemble of distributed agents, with the aim of letting them reach an agreement about their beliefs and/or plans of actions [37]. In the speaking object scenario, conversations take place by having speaking and hearing objects exchange *assertions* about the current or desirable state of the affairs, respectively. Such assertions can be contradicted or strengthened by others engaging in the conversation with the goal of reaching an agreement about the state of the world (for speaking objects) or about a joint plan aimed at achieving a given state of affairs (for hearing objects).

Conversational approaches to distributed coordination are radically different from traditional approaches, which tend to enforce strict rules on the behavior of components, and assume the presence of specific *coordination laws* to respect, in terms of

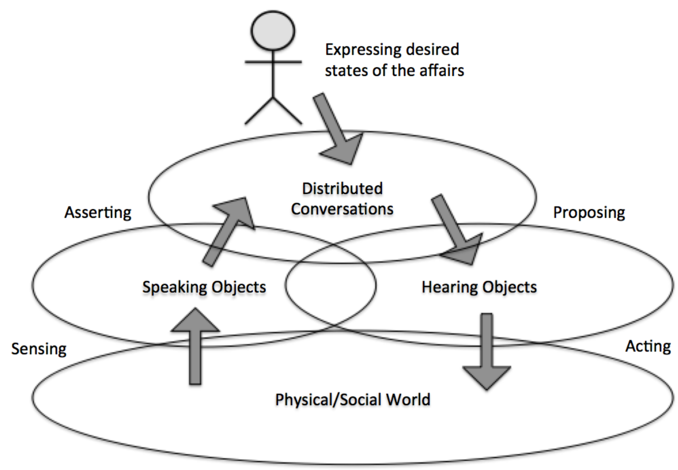

Fig. 2. Coordination among smart speaking objects and smart hearing actuators has to be realized as a sort of distributed multi-agent conversation.

how components interact and how components should behave during interaction. They mostly leave no room for goal-oriented behaviors and for *adapting* the dynamics of a distributed coordination protocol to the actual outcomes of the conversation itself and to the arguments raised by components during the coordination process.

In some sense, conversation-based coordination shifts attention to the *meta-level* of coordination, by providing rules to negotiate interaction protocols rather than the protocols themselves. *Flexibility* greatly benefits from this perspective, because not only the actual interactions among participant components arise at run-time according to a given interaction protocol, but the protocol itself *emerges* from the bottom up. Furthermore, traditional coordination approaches are mostly *memoryless*, as they rarely track the history of interactions for purposes beyond performance tuning, computation of trust, or adaptation of policies. The envisioned conversations, instead, naturally account for interaction history through the notion of *commitment*, aiming to track promises, claims, and arguments, for the sake of correctness of the whole coordination process.

Even in the IoT arena, most approaches for orchestrating the activities of the different components rely, as of today, on a set of *rules*, and on middleware engines that check and enact them [33]. Such rules dictate how the components should be activated (and their services executed), depending both on the situations that are happening, and on those that – in reaction – should be achieved. However, in a scenario of speaking and hearing (goal-oriented) objects, such an approach falls short, due to the impossibility of foreseeing and defining all possible events and state of the affairs, and all the possible ways in which components can be activated. It is in fact unfeasible to design all the possible composition rules that orchestrate the behaviors of the components. Thus, while the possibility of defining rules and constraints for the “do” and the “don’t” of the systems (e.g., *safety* and *liveness* properties that should be always guaranteed [42]), the actual way the components act and interact should be identified at run-time by the components themselves, still in respect of global system goals and constraints.

The issue of reaching a consensus in an ensemble of interacting autonomous components via distributed negotiations has been deeply investigated in the area of *agent-oriented* computing [18]. However, negotiation mechanisms are blind with respect to the strategy adopted by the agents participating in the negotiation. This does not help in reaching globally satisfactory solutions, which could be achieved instead by letting agents conversate and motivate their choices, as proposed in *argumentation-based* multi-agent negotiation [31], a research area that has very strong relations with our vision (see Section 4.4).

3.2 Types of Conversations

Let us now classify the different types of conversation that one can expect to take place in the speaking objects scenario.

Among Speaking Objects. Speaking objects are likely to interact with each other in order to build and report a complete and coherent understanding of their surroundings. However, it may be the case that the identification of a specific situation requires *(i)* more information than initially thought, or *(ii)* solving some conflicting perceptions.

The former case triggers what are called *information seeking* and *inquiry dialogues* [37]. These are aimed at integrating the originally incomplete information with either new information or more arguments in support of the existing information. For example, in the smart hospital scenario, a set of speaking cameras need to ask each other who they are detecting to collectively build a global map of patients' locations in real-time.

In the latter case, different (sets of) speaking objects may reach different conclusions about what is happening, which triggers *negotiation* and *persuasion dialogues* to let them all agree on a common perspective. To this end, speaking objects may exchange arguments explaining the reasons why they ended up identifying a specific situation to persuade others, or they may decide to involve additional sensors in the conversation. In a smart hospital scenario, the variety of speaking objects may not necessarily acquire the same perspective on what is sensed. A camera in the rehabilitation room of the hospital may recognize that a man is "running on the tapis roulant", the tapis roulant itself may state that the user is "standing", whereas the wristband may recognize that he is "jumping". To solve the conflict, they may start comparing with each other the reasons behind their respective understandings of the situation. This can enable discovering that, since the tapis roulant is off (and this is why it stated that the user was "standing"), the only reasonable explanation is that "the user is jumping on the tapis roulant".

We emphasize that, although a variety of sensor fusion techniques exist to support situation identification [24], these typically act downstream the sensor level, as they simply receive data from sensors and try to apply well-defined rules to both integrate distinct data streams and solve possible conflicts. Basically, they are mostly *black-boxes* from an observer standpoint. Moreover, they do not usually consider giving sensors the possibility of taking action themselves. Yet, in our view speaking sensor objects become sort of *grey-boxes*: they can be requested to *justify* their perceptions and *explain* their course of action, and are expected to provide insights into the reasoning that guides their behavior. The same holds for hearing actuator objects, as described in the following.

Between Speaking and Hearing Objects. While planning for a specific course of action aimed at achieving a given state of the affairs, hearing objects may recognize that they need more information and/or more convincing arguments than initially provided in order to make an informed decision.

This kind of conversation is a mixture of *information seeking*, *inquiry*, and *deliberation* dialogues [37], which should be suitably composed so as to enable informed decision making: in this way, hearing actuators are able to plan and justify their course of actions based on the amount and quality of information required by the scenario at hand. Notice that this kind of close *feedback loop* between sensing and acting is very expensive with state of the art cloud-based approach to IoT.

Among Hearing Objects. In the majority of real world applications, such as in the traffic management and assisted living scenarios already described, it is quite unusual that actuators are able to *individually* change their environment (namely, act) so as to achieve the optimal state of affairs. Rather, it is usually through collaboration and joint planning efforts that the most effective and efficient strategy to achieve a given goal can be designed and pursued. Accordingly, it is often the case that hearing objects engage in *deliberation* dialogues, meant to achieve a *shared plan* by exchanging arguments about feasibility of actions, their expected utility, likelihood of positive/negative outcomes, and the like. Then, it is similarly unrealistic to assume that the landscape of all the possible actions by all the participant actuators is *conflict-free* [43]. Thus, *negotiation* and *persuasion* dialogues are required as a means to argue toward conflicts resolution.

As an example, consider an A/C system in a room of the hospital willing to turn itself on after hearing the thermostat assert "it's hot". In case a few hearing windows are also installed, both the A/C and the windows may decide to act, without actually generating any conflict: either turning on the A/C or opening the windows (or doing both) leads to the goal anyway. Nevertheless, doing both is sub-optimal from the standpoint of efficiency, thus joint deliberation to *collectively* choose an individual course of action or a shared plan – in this case, who acts and who doesn't – is likely welcome. Accordingly, the window may convince the A/C not to act by argumenting "there is a fresh breeze outside, I can save power consumption while still chilling the room". Now consider the same scenario during the summer: if both actuators act there is a conflict, because the air coming from the outside would likely be hot, actually neglecting the air conditioning effect—or, at the very least, hindering the A/C system course of actions and leading to sub-optimal efficiency and effectiveness. Yet again, thus, joint deliberation for shared planning is required.

4 Enabling Technologies

Let us now present the main technologies and approaches which enable our vision. Although these have been widely investigated in the context of agents and multiagent systems, they are not (yet) properly accounted for by research in the IoT area.

4.1 Cognitive Reasoning

A first cornerstone of the architecture that we propose is certainly a framework for cognitive reasoning by speaking objects, in particular grounded on *knowledge representation* and *commonsense reasoning*.

Speaking and hearing objects will continuously undertake conversations among them and with humans, and thus they will need to share a coherent representation of the world. Large-scale ontologies will be needed as repositories to interpret the knowledge bases available to the agents. Such knowledge could be continuously modified, adapted and refined by the agents themselves, according to their experience and perception of the environment. This aspect, although widely explored in the area of intelligent agents, is still somehow under-explored in the IoT domain [15]. In fact, although advanced classification and recognition techniques are widely employed in IoT scenarios, they are mostly tailored to specific application domains or business goals, such as activity recognition [39]. Both a general model and a supporting infrastructure for *context understanding*, *knowledge inference*, and *dynamic composition* are still missing.

Some recent studies tried to design architectures capable of addressing these tasks, but the general problem is certainly far from being solved. Ganz et al. [12], for example, distinguish lower-level and higher-level abstractions, as granules of information that are constructed starting from raw data. Lower-level abstractions represent "static and atomic information" that can be extracted from a single sensor (e.g., a light is on or off), whereas higher-level abstractions can be inferred from multiple lower-level abstractions to construct more complex patterns (e.g., from the facts that the light in the kitchen is on, it is 7 p.m., and the oven has just been switched on, one could infer that someone is cooking). The authors sketch some possible research directions within this context, by highlighting some existing off-the-shelf machine learning and data mining systems that could be exploited in the scenario. Perera et al. [30] present an extensive survey on what they name "Context Aware Computing" in the IoT domain. They revise and compare many different pervasive computing systems that operate in the IoT setting, by highlighting their pros and cons, and by focusing on whether and how they are able to incorporate background knowledge and context-awareness. From their careful analysis, it is evident that most of the existing systems adopt a *rule-based* mechanism to model background knowledge and thus they perform reasoning at a symbolic level, often building upon existing ontologies. Clearly, such an approach relies on handcrafted knowledge bases, specifically designed by experts for a particular domain.

On a very similar perspective, commonsense reasoning is another key ingredient of the framework. It refers to a line of research that tries to give computers the ability to make presumptions about ordinary situations that humans encounter every day. These assumptions include judgments about the physical properties, purpose, intentions and behavior of people and objects, as well as possible outcomes of their actions and interactions [9]. For instance, if an autonomous vehicle sees a ball rolling across the street, it can assume a kid is running after it, thus decide to stop accordingly. As another example, in a smart home scenario where a six-foot-tall person and a two-foot-tall person are at home, if the system is instructed to look for the "toddler", it does not have to ask which is which. The more computers will be integrated in our everyday life, the more this kind of knowledge will be fundamental to allow them to operate autonomously and proactively. However, the current state of the art is very limited: existing proposals can cope with restricted domains (and in general it is very difficult to understand how much commonsense knowledge has been covered), and with restricted reasoning capabilities (typically, taxonomic reasoning). Research is likely to improve representation and reasoning capabilities together, given the increasing availability of large amounts of data about our everyday life in multiple domains.

4.2 Machine Learning

Massively distributed sensors in the IoT arena clearly produce huge data streams, that need manipulation, aggregation, and sometimes also more sophisticated, intelligent elaboration. These steps are nowadays tipically performed directly on-board, within smart sensors, that more and more commonly by now contain tools such as deep networks [22]. Turning the processed information into high-level knowledge is, however, still an open issue [30].

Another peculiar trait of speaking and hearing objects is the capability of learning behaviors, strategies, and policies from historical data and situations, with the aim of continuously adapting to the environment. This would represent a major advantage with respect to approaches based on sets of pre-defined, hand-crafted rules, that are clearly hard to update in case of abrupt system changes. Similarly, pattern mining methodologies could be exploited to perform association rule mining and user profiling [36]. Here, we believe that Statistical Relational Learning [14] and Neural-Symbolic learning [13] could offer a valuable research direction to pursue, as they propose to combine logic-based approaches with statistical learning, probabilistic models and neural approaches (including deep learning), with the goal of both handling uncertainty in data, and exploiting background knowledge. The idea is that grey-box models, capable of exploiting both the computational power of systems such as deep networks, and the interpretability of logic and argumentation, will offer tools to support medium and long-term self-adaptation of pervasive computing systems.

Being dialogue and coordination another crucial ingredient of speaking and hearing objects, the long-standing research field of learning in multi-agent systems [35] will need to be integrated in this conversational context. Coordination, in fact, will have to be performed through natural language, in order to easily insert humans in the loop, and to offer interpretable explanations of the dynamics of the system.

Finally, also actuators will be more deeply rooted with machine learning technologies. Although many IoT actuators provide just a simple interface, with a limited number of possible actions, however there are many contexts where complex behaviors are needed. Robots and self-driving cars, and their components, are an example of such a complex, adaptive actuator. Such devices are typically told just a goal to achieve (such as reaching a destination), without explicitly detailing every step to take, which the devices have to find out themselves, possibly through coordinaton, as it happens, for example, in the area of swarm robotics [5].

4.3 Goal-oriented Computing

Making actuators become goal-oriented requires to ascribe them a few crucial capabilities: *(i)* recognize expression of a goal, as a state of affairs to be achieved; *(ii) deliberate* whether they may play a role in pursuing that goal, and how; *(iii) reason* about feasibility, likelihood of success, and outcomes of the actions needed to get there [38]; *(iv) plan* the course of actions to undertake, considering cost, expected utility, etc. [28]. All of this in *autonomy*, that is, with the opportunity to reject goals if they are not of interest, abandon them if they are no longer feasible, offer help to others if a collaboration opportunity arise, as well as ask help to others if no means to achieve the goal are currently possessed by the actuator.

It is worth noting that goal-oriented behaviour may be ascribed to speaking objects as well. In the current IoT vision, sensors are simply *hard-coded* to monitor a given property of a given environment, to generate data and events accordingly (e.g., a thermostat sensing the temperature each second, and reporting each time a "new temperature value" event). In the speaking objects vision, instead, sensors may bind monitoring activities to an *explicit* and *dynamic* goal, either expressed by another component or by a human user.

It is then necessary to embed at the very foundation of the speaking objects vision all the concepts, abstractions, and models commonly found in the *agent-oriented* literature, such as the notion of cognitive agents [32], techniques for means-ends reasoning [38] and planning [28], the many issues of coordination in multi-agent systems [29]. Many languages and infrastructures have proven to be mature enough for relevant scenarios in the agent-based community: for a survey, the interested reader is referred to [4]. Yet, their viability and effectiveness in a highly dynamic, heterogeneous, resource-constrained, and scale-demanding domain such as IoT, still remains to be fully assessed.

4.4 Argumentation-based Coordination

Argumentation is required as a necessary feature of sensor and actuator devices to regard them as speaking and hearing objects. Argumentation may in fact well support: *(i) decentralised coordination*, by leveraging negotiation opportunities; *(ii) situated reasoning*, by enabling belief revision in face of uncertainty; *(iii) joint deliberation*, by allowing negotiation over desires and plans besides beliefs; *(iv) "humans-in-the-loop"*, by making explanations and justifications of decision making available in natural language. For a more thorough analysis of these aspects, the reader may refer to [25].

Despite the long history of research in argumentation, only recently practical applications to real-world scenarios have started receiving attention (e.g., see [20]). Furthermore, for argumentation to work there must be either an agreement among participants about the *admissible moves* and their significance, or an *external judge* enacting some form of control over the argumentation process. Neither of the two is straightforward to have in the speaking objects vision: reaching agreement is difficult *per se*, besides being unlikely easily scalable; and having an external authority may be an unacceptable centralisation point.

A way out can be found by carefully investigating *hybrid* approaches where, for instance, a multitude of external authorities share the load of arbitrating argumentations among a limited number of participants, possibly exploiting some notion of physical or logical proximity to enforce shared argumentation rules. Another solution could be to have participants agree only temporarily, for the duration of a given "conversation session" on a common set of argumentation rules, which may then change for future conversations depending on, e.g., timing constraints or the type of dialogue.

4.5 Human-Computer Interaction

Smart objects of the future will certainly need to send and receive commands exploiting natural interfaces, such as voice commands or gestures. Techniques coming from natural language processing, speech recognition, computer vision will become essential components of smart objects in the near future, as they already are in our smartphones. In this way, less effort will be required to program devices, and users will experience a more direct and *transparent* interaction with technology [23].

Recent advances in machine learning are finally realizing the vision of conversational interfaces in the form of natural language *chatbots*. On this basis, a number of companies are pushing the idea of interacting with the IoT devices via chatbots (e.g., Amazon' Echo and Google Home) both to give instructions to devices and to let them report and give information to users. While the current state of the art is about interacting with a single device or hub, in the near future we envision interacting with *many* at the same time. For example, a voice command will be heard by multiple devices, and each will have to interpret it, as well as to understand its role in the overall fulfillment. Similarly, when reporting about a situation, IoT devices will need to coordinate a coherent story to be told to the user. As a further example, let us imagine a user entering a smart home and asking "Why it is so cold?". The question will be heard by multiple devices, most of which will not have any role about that situation and thus will disregard it. Instead, other devices such as (smart) HVAC, thermostat, windows, etc. will recognize their potential involvement and start interacting to come up with a possible explanation and solution: by exploiting commonsense reasoning (see Section 4.1) they will realize that the user is not only asking for information, but for a solution as well.

5 Integration Recipe: Open Challenges for Realizing the Vision

Although we identified some technologies that will most likely become key ingredients in the speaking object vision, actually realizing the vision implies having the appropriate modelling tools and middleware infrastructures to coherently integrate them, and to ensure they will be employed to produce practical, usable, and dependable systems.

5.1 Massive Scale and Heterogeneity

The key challenge in developing and controlling systems of distributed speaking objects is their massive overall scale. It is foreseen that in the near future billions of IoT devices will populate our cities, including thousands of our buildings and homes. Such myriads of devices will be in need to be coordinated at different scales, from the global ones (e.g., for achieving policies at urban level) to the local ones (i.e., for realizing functionalities and achieving policies at building or home level). The computational power of these smart devices is growing faster and faster, allowing to embed very advanced technologies in relatively cheap hardware. This will be a key factor for a massive distribution of intelligent, autonomus agents.

Accordingly, on the one hand there will be need to design and deploy coordination schemes that can support coordination among a very large number of distributed components, to realize global policies. However, these can hardly rely on conversations and argumentation-based approaches, whose scalability remains an open issue. Rather, they should get inspiration from social and nature-inspired coordination models [41].

On the other hand, the above forms of large-scale coordination should co-exist with more local, argumentation-based, forms of coordination to achieve local goals. How the two forms of coordination could co-exist is definitely an open and fascinating research challenge.

5.2 Middleware

From a more implementation-oriented perspective, given that conversations are a new means of coordinating the activities of distributed components, a key open research question is to understand what services should a middleware provide to support such distributed conversations.

Although conversation can be assimilated to a sort of message-passing interaction, a mere message-oriented middleware (MOM) would not be enough [7]. In fact, the course of a conversation implies the existence and the creation of shared knowledge among the interacting components, which cooperatively build a shared perspective of the world based on *logically sound* and connected arguments, and cooperatively build a shared plan of actions. MOM also falls short in supporting interaction in a dynamic world, where the identities and characteristics of components are not known in advance, as in the case of speaking objects (and of IoT in general).

Accordingly, the coordination model supported by the middleware should not simply provide the functionality of MOMs in terms of direct interactions between components. Rather, it should support conversations via an open and shared conversation space, enabling conversation among components that do not necessarily have to know each other in advance, and also providing services for sharing knowledge and information, e.g., a tuple space [11]. However, unlike traditional tuple space models, which contain unrelated chunks of information, the need to access information about conversations implies connecting information into sorts of *knowledge networks* detailing how conversations evolved and how they are related. Although some proposals in that direction exist [27], the best way to realize such shared conversation space will have to be evaluated. As it is yet to be evaluated how corpora of commonsense knowledge could be integrated within the overall architecture to support conversations.

5.3 Humans-in-the-Loop

The speaking objects vision cannot overlook *humans-in-the-loop* as a vital computational component of the scenario. In fact, besides participating as actors that impose their desired states of the affairs to the system (see Figure 2), humans can become actual components of the system itself: they can participate by providing sensing capabilities (thus acting as speaking objects), actuating capabilities (as hearing objects), and can consequently be involved in conversations. This convergence between human and software entities is witnessed by many modern *socio-technical systems*, and it demands researchers and practitioners to conceive, design, and develop systems seamlessly interacting with other software systems and with human agents as well.

It is worth noting that when human users enter the picture, the need for argumentation-based conversations is even more evident: the ability of smart objects to justify their stances, in fact, it becomes crucial to convince users to effectively participate in the conversational process. Clearly, this may require accounting for socio-cognitive models of action and interaction as they can be observed among human agents, to be suitably transferred to the synthetic domain of conversating speaking objects.

Besides the need for effective means of human-machine interaction, as already discussed in Section 4, integrating humans in the loop also challenges the whole software engineering process, the modeling and design of human behaviours and of conversations involving humans, and the functionalities that the middleware should provide to enable integration.

5.4 Harnessing Algocracy

We are already, for some aspects of our life, living in an *algocracy* [8]: we are passive subjects of choices automatically made by some algorithms on our behalf. The most typical example being the filtering that social networks perform in presenting us their news feed. The problem of being passive subjects to some form of algorithmic governance, in the era of, e.g., smart cities and self-driving cars, will be increasingly relevant, because it will start affecting each and every moment of our lives. In the scenario of speaking objects, the way our everyday objects and environment will behave will ultimately depend on a variety of arguments, conversations, and joint plans of actions, possibly including the resolution of conflicting goals.

For this reason, it is of paramount importance that speaking and hearing objects assume the form of "grey-boxes": not only to enable understanding each other arguments, but also to enable humans whose lives are affected by their decisions and arguments to understand what happens, and possibly to intervene as needed. One can see this as a different facet of the previously discussed "humans in the loop" issue. Not only humans should be able to interact with the system to provide goals and to act as human sensors or actuators. They must be enabled to interact with the system to *inspect* its way of acting and understand how the world around them works.

This raises another challenge related to *algorithmic literacy*. As we need at least some (even intuitive) knowledge of physics to interact with the physical world, we might need some form of literacy about how IT components perceive, interact, and deliberate actions in a cyberphysical world. And this will be a different kind of – and yet to be invented – literacy than that currently promoted by "learning to code" campaigns.

5.5 Security

Security is and will likely always be a major concern for IoT systems and applications [19], involving critical safety issues when applied to, e.g., smart assistance scenarios.

It is clear that the deployment of speaking objects systems will have to deal with security and safety as well. Indeed, the presence of humans participating in conversations with (grey-box) speaking and hearing objects, with the possibility of somehow directing the outcomes of such conversations, can represent a door for attacking the system, stealing sensitive information, and introduce (possibly safety critical) malfunctioning.

At the same time, the capability of speaking objects of argumenting about situations, can potentially represent also a mean to protect from attacks. By making speaking objects able to understand the reasons behind the evolution of a conversation and of the overall behavior of the system, objects could be made somewhat capable of recognizing malicious behaviours and threats. In other words, by leveraging over existing methods on argumentation-based risk assessment [16], automated argumentation-based security could become a thing.

6 Conclusions

The emergence of speaking objects will dramatically change the approaches to implementing and coordinating the activities of distributed IoT processes and services, calling from bringing in the lessons of multiagent systems, requiring to face scalability issues, and requiring integration of a number of technologies and approaches from different areas that are likely to play a key role towards the envisioned scenario.

In any case, for the vision to turn into a practical and useful reality, we have identified a number of research challenges to be further investigated. These, and possibly further issues we may have failed in identifying, can represent a fascinating playground for researchers in the area of distributed and pervasive computing systems.

Acknowledgments: Work supported by the CONNECARE (Personalised Connected Care for Complex Chronic Patients) project (EU H2020-RIA, Contract No. 689802).

References

[1] H. Agrawal, S.-w. Leigh, and P. Maes. L'evolved: Autonomous and ubiquitous utilities as smart agents. In *ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 487–491, New York, NY, USA, 2015. ACM.

[2] L. Amgoud and S. Parsons. *Agent Dialogues with Conflicting Preferences*, pages 190–205. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.

[3] L. Atzori, A. Iera, and G. Morabito. The internet of things: A survey. *Computer Networks*, 54(15):2787–2805, 2010.

[4] R. Bordini, L. Braubach, M. Dastani, A. El Fallah Seghrouchni, J. Gomez-Sanz, J. Leite, G. O'Hare, A. Pokahr, and A. Ricci. A survey of programming languages and platforms for multi-agent systems. *Informatica (Ljubljana)*, 30(1):33–44, 2006. cited By 152.

[5] M. Brambilla, E. Ferrante, M. Birattari, and M. Dorigo. Swarm robotics: a review from the swarm engineering perspective. *Swarm Intelligence*, 7(1):1–41, 2013.
[6] M. Conti, S. Das, C. Bisdikian, M. Kumar, L. Ni, A. Passarella, G. Roussos, G. Troster, G. Tsudik, and F. Zambonelli. Looking ahead in pervasive computing: Challenges and opportunities in the era of cyber-physical convergence. *Pervasive and Mobile Computing*, 8(1):2 – 21, 2012.
[7] E. Curry. Message-oriented middleware. *Middleware for communications*, pages 1–28, 2004.
[8] J. Danaher. The threat of algocracy: Reality, resistance and accommodation. *Philosophy & Technology*, 29(3):245–268, 2016.
[9] E. Davis and G. Marcus. Commonsense reasoning and commonsense knowledge in artificial intelligence. *Communications of the ACM*, 58(9):92–105, 2015.
[10] M. Endler, J.-P. Briot, F. Silva e Silva, V. P. De Almeida, and E. H. Haeusler. An Approach for Real-time Stream Reasoning for the Internet of Things. In *Proceedings of the 11th IEEE International Conference on Semantic Computing (ICSC'2017)*, pages 348–353, San Diego, Californie, United States, Jan. 2017. IEEE, IEEE.
[11] P. T. Eugster, P. Felber, R. Guerraoui, and A. Kermarrec. The many faces of publish/subscribe. *ACM Comput. Surv.*, 35(2):114–131, 2003.
[12] F. Ganz, D. Puschmann, P. Barnaghi, and F. Carrez. A practical evaluation of information processing and abstraction techniques for the internet of things. *IEEE Internet of Things journal*, 2(4):340–354, 2015.
[13] A. S. d. Garcez, K. B. Broda, and D. M. Gabbay. *Neural-symbolic learning systems: foundations and applications*. Springer Science & Business Media, 2012.
[14] L. Getoor and B. Taskar. *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2007.
[15] A. Gyrard, M. Serrano, and G. A. Atemezing. Semantic web methodologies, best practices and ontology engineering applied to internet of things. In *Internet of Things (WF-IoT), 2015 IEEE 2nd World Forum on*, pages 412–417. IEEE, 2015.
[16] S. Ingolfo, A. Siena, J. Mylopoulos, A. Susi, and A. Perini. Arguing regulatory compliance of software requirements. *Data & Knowledge Engineering*, 87:279–296, 2013.
[17] S. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak. The internet of things for health care: a comprehensive survey. *IEEE Access*, 3:678–708, 2015.
[18] N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, M. J. Wooldridge, and C. Sierra. Automated negotiation: prospects, methods and challenges. *Group Decision and Negotiation*, 10(2):199–215, 2001.
[19] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu. Security of the internet of things: perspectives and challenges. *Wireless Networks*, 20(8):2481–2501, 2014.
[20] H. Jung, M. Tambe, and S. Kulkarni. Argumentation as distributed constraint satisfaction: Applications and results. In *Proceedings of the Fifth International Conference on Autonomous Agents*, AGENTS '01, pages 324–331, New York, NY, USA, 2001. ACM.
[21] J. O. Kephart and D. M. Chess. The vision of autonomic computing. *IEEE Computer*, 36(1):41–50, Jan. 2003.

[22] G. Kortuem, F. Kawsar, V. Sundramoorthy, and D. Fitton. Smart objects as building blocks for the internet of things. *IEEE Internet Computing*, 14(1):44–51, 2010.
[23] M. Kranz, P. Holleis, and A. Schmidt. Embedded interaction: Interacting with the internet of things. *IEEE internet computing*, 14(2):46–53, 2010.
[24] M. Liggins II, D. Hall, and J. Llinas. *Handbook of multisensor data fusion: theory and practice*. CRC press, 2017.
[25] M. Lippi, M. Mamei, S. Mariani, and F. Zambonelli. An argumentation-based perspective over the social iot. *IEEE Internet of Things Journal*, pages 1–1, 2017.
[26] M. Lippi, M. Mamei, S. Mariani, and F. Zambonelli. Coordinating distributed speaking objects. In *37th IEEE International Conference on Distributed Computing Systems, ICDCS 2017, Atlanta, USA, June 5-8*, 2017.
[27] S. Mariani. *Coordination of Complex Sociotechnical Systems - Self-organisation of Knowledge in MoK*. Artificial Intelligence: Foundations, Theory, and Algorithms. Springer, 2016.
[28] F. R. Meneguzzi, A. F. Zorzo, and M. da Costa Móra. Propositional planning in bdi agents. In *Proceedings of the 2004 ACM Symposium on Applied Computing*, SAC '04, pages 58–63, New York, NY, USA, 2004. ACM.
[29] A. Omicini and M. Viroli. Coordination models and languages: From parallel computing to self-organisation. *The Knowledge Engineering Review*, 26(1):53–59, Mar. 2011.
[30] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos. Context aware computing for the internet of things: A survey. *Communications Surveys Tutorials, IEEE*, 16(1):414–454, First 2014.
[31] I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. Mcburney, S. Parsons, and L. Sonenberg. Argumentation-based negotiation. *Knowl. Eng. Rev.*, 18(4):343–375, Dec. 2003.
[32] A. S. Rao and M. P. Georgeff. BDI agents: From theory to practice. In V. R. Lesser and L. Gasser, editors, *1st International Conference on Multi Agent Systems (ICMAS 1995)*, pages 312–319, San Francisco, CA, USA, 12-14 June 1995. The MIT Press.
[33] M. A. Razzaque, M. Milojevic-Jevric, A. Palade, and S. Clarke. Middleware for internet of things: A survey. *IEEE Internet of Things Journal*, 3(1):70–95, Feb 2016.
[34] W. Shi and S. Dustdar. The promise of edge computing. *Computer*, 49(5):78–81, May 2016.
[35] P. Stone and M. M. Veloso. Multiagent systems: A survey from a machine learning perspective. *Auton. Robots*, 8(3):345–383, 2000.
[36] C.-W. Tsai, C.-F. Lai, M.-C. Chiang, L. T. Yang, et al. Data mining for internet of things: A survey. *IEEE Communications Surveys and Tutorials*, 16(1):77–97, 2014.
[37] D. Walton and E. Krabbe. *Commitment in Dialogue: Basic concept of interpersonal reasoning*. State University of New York Press, Albany NY, 1995.
[38] M. J. Wooldridge. *Reasoning about rational agents*. MIT press, 2000.
[39] J. Ye, S. Dobson, and S. McKeever. Situation identification techniques in pervasive computing: A review. *Pervasive and Mobile Computing*, 8(1):36–66, 2012.

[40] S. Yi, C. Li, and Q. Li. A survey of fog computing: Concepts, applications and issues. In *Proceedings of the 2015 Workshop on Mobile Big Data*, Mobidata '15, pages 37–42, New York, NY, USA, 2015. ACM.
[41] F. Zambonelli and et al. Developing pervasive multi-agent systems with nature-inspired coordination. *Pervasive and Mobile Computing*, 17, Part B:236–252, 2015.
[42] F. Zambonelli, N. R. Jennings, and M. Wooldridge. Developing multiagent systems: The gaia methodology. *ACM Trans. Softw. Eng. Methodol.*, 12(3):317–370, 2003.
[43] M. R. Zatelli, J. F. Hübner, A. Ricci, and R. H. Bordini. Conflicting goals in agent-oriented programming. In *Proceedings of the 6th International Workshop on Programming Based on Actors, Agents, and Decentralized Control, AGERE*, pages 21–30, 2016.