



# Distributed Speaking Objects: A Case for Massive Multiagent Systems

Marco Lippi<sup>(✉)</sup> , Marco Mamei , Stefano Mariani ,  
and Franco Zambonelli 

Dipartimento di Scienze e Metodi dell'Ingegneria,  
Università di Modena e Reggio Emilia, Reggio Emilia, Italy  
{marco.lippi,marco.mamei,stefano.mariani,franco.zambonelli}@unimore.it

**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 arguing 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 objects vision, overview the key enabling technologies, and analyze the key challenges for engineering large-scale collectives of speaking objects and their conversations.

**Keywords:** Massive multiagent systems · Internet of Things · Argumentation

## 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 [5].

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*, often cloud-based 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 [1,9,34] let us point to a novel scenario:

- IoT devices can and are going to become much smarter [9]. 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” [24].
- 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 local 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 *arguing* about situations and about the current “state of the affairs” [9], 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.

In this context, the paper provides the following contributions:

- An analysis of the key concepts behind speaking objects, showing how they are going to change the very nature of decentralized coordination and are going to challenge traditional approaches to distributed computing and calling for novel *conversational* approaches.
- An overview of 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.
- The identification of some 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 exemplary case study, we will consider the case of a large-scale deployment where a *smart hospital* is instrumented to support

health monitoring and assisted living [16]. 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., spirometers, heartbeat monitoring devices, 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

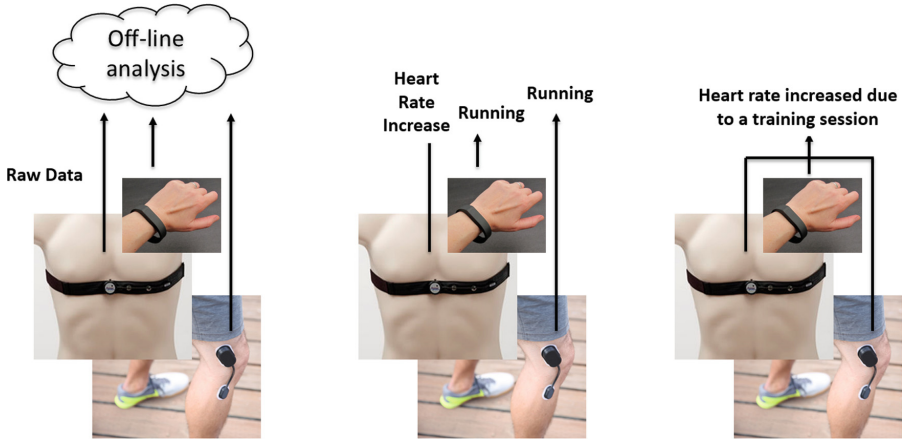
Currently, in the IoT arena (and in related typical application scenarios, from smart homes to smart cities and transportation) the concept of *smart object* is mostly associated to the possibility of attaching ICT devices to physical objects and places, thus turning them into: (i) *sensors*, capable 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 [39]) point for later analysis; (ii) remotely controllable *actuators*, capable of enacting specific configurations or actions in the surrounding environment, by receiving appropriate commands.

Progress across many different areas, though, 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 Data Collection vs. Cognitive Sensing

Advancements in machine learning techniques, and in the increase of computational power that can be embedded in everyday sensors and objects, is making it possible for smart objects to analyze *locally* the stream of sensed data in order to extract relevant features from it. A simple example, in our case study scenario, is a set of wearable devices monitoring physiological parameters and physical activities of a patient, capable of associating the sensed patterns of movement to situations like “unusual heart rate”, “walking”, “running” (see Fig. 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 set of wearables may construct the assertion that “Heart rate increased due to a training session” from the sensing of two distinct patterns. Or a camera may perform scene understanding, by



**Fig. 1.** From simple sensors to speaking objects. In a smart hospital scenario a number of wearable devices can interact – speak – to gather a complete description of a situation.

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 [38].

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.

In any case, 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 which they are delivered with.

## 2.2 Actuating Commands vs. Achieving Goals

Concerning 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, adjusting lightning accordingly. Or, a smart desk lamp that autonomously moves and tunes intensity to ensure optimal illumination in spite of changing environmental conditions [1].

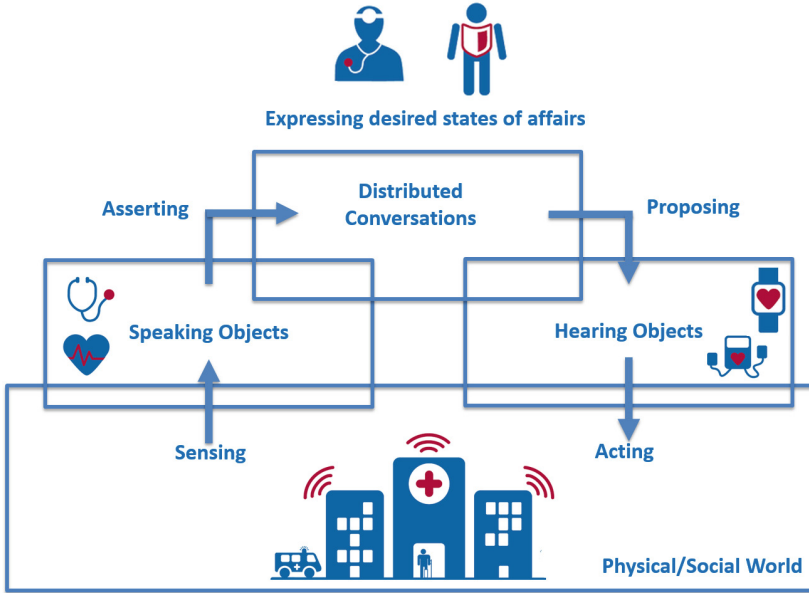
Smart actuator objects, to achieve their goals, must acquire information about the current state of the affairs, which requires gathering information from smart sensors. Also, they must sometimes interact with each other and with non-smart objects (e.g., non goal-oriented actuators). For instance, in 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 Fig. 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 to 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* [19], 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 [39].



**Fig. 2.** Coordination among smart speaking objects and smart hearing actuators has to be realized as a sort of distributed multi-agent conversation. In the smart hospital scenario a massive amount of devices and systems might need to coordinate to obtain a coherent view of the situation.

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 [36]. 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 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 [32]. 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 [40]) should remain, 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 [17]. 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 converse and motivate their choices, as proposed in *argumentation-based* multi-agent negotiation [30], a research area that has very strong relations with our vision (see Sect. 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* [36]. These are aimed at integrating the originally incomplete information with either new information or more arguments in support of the existing one. 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 the 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 treadmill”, the treadmill 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 treadmill 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 treadmill”.

We emphasize that, although a variety of sensor fusion techniques exist to support situation identification [22], 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 [36], 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 closed *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 assisted living scenario 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

First of all, given their conversational nature, speaking and hearing objects need to implement some form of cognitive reasoning, and especially of *knowledge representation* and *commonsense reasoning*. By continuously interacting among them and with humans through dialogue, they will have to share a common representation of the world.

A clear need is that of exploiting knowledge bases and large-scale ontologies to model and represent the concepts and their relations, which the agents continuously deal with. This issue represents a significant challenge in agent coordination [10] and it remains under-explored in the IoT domain [14]. Although the general problem is far from being solved, yet some recent works have proposed architectures that address the aforementioned issues. For example, in [11]

a framework is proposed, that builds lower- and higher-level abstractions, starting from raw data. A recent survey [29] presents several approaches to context-aware computing in the IoT domain, with a specific emphasis on their capability to embed background knowledge and context-awareness. Such thorough analysis shows how *rule-based* mechanisms are still largely employed to perform symbolic reasoning, thanks to the hand-crafted knowledge bases designed by experts. An analysis of the scalability of this kind of technologies towards massive systems has been recently presented [25], together with an experimental evaluation of the most promising semantic reasoning approaches in the IoT arena.

Commonsense reasoning also has to be integrated into the scenario of speaking and hearing objects. This keyword describes a research area where the aim is to make computers capable of performing those basic inference processes that we, as humans, continuously perform without even thinking [8]. This skill is crucial in our everyday life, and allows us to take decisions and solve problems. Smart devices that will be more and more integrated in our life, such as speaking and hearing objects, will necessarily embed this ability in order to autonomously and proactively operate. Currently, existing approaches are limited to restricted domains and, therefore, to restricted reasoning capabilities (typically, taxonomic reasoning) [8]. We argue that large-scale scenarios will provide novel data collections upon which it will be possible to test new techniques, for example coming from machine learning.

## 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 often performed directly on-board, within smart sensors, that can embed tools such as deep networks [20]. Turning the processed information into high-level knowledge is, however, still an open issue [29].

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 [35]. Here, we believe that Statistical Relational Learning [13] and Neural-Symbolic learning [12] 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. In this way, speak-

ing objects will move a step towards *explainable artificial intelligence*, which is considered one of the major challenges for the near future.

### 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 [37]; *(iv)* *plan* the course of actions to undertake, considering cost, expected utility, etc. [27]. 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 such an opportunity arises, and ask help to others if no other means to achieve the goal is currently available.

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. 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 [31], techniques for means-ends reasoning [37] and planning [27], the many issues of coordination in multi-agent systems [28]. 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 [23].

Despite the long history of research in argumentation, only recently practical applications to real-world scenarios have started receiving attention (e.g., see [18]). 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.

## 5 Integration Recipe: Open Challenges for Realizing the Vision

Although we identified some technologies that will most likely become key ingredients in the speaking objects 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. In fact, this enables efficient *separation of concerns*, that is distributing functionalities and responsibilities, among the different scales of the system, so as to better tackle the most pressing issues at the right level of abstraction: for instance, critical functionalities requiring rapid decision making and adaptation for quickly solving local contingencies can be attributed to the smaller scale of the multi-scale system at hand (such as an hospital), up to the individual device, whereas medium and long term planning and scheduling of strategic actions can be charged upon the higher scales of the system (i.e., a department-wide in-house server scheduling appointments, or a hospital-wide cloud-based platform planning resource exploitation).

Accordingly, on the one hand it will be needed 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 [42]. 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.

In the case of the hospital deployment already mentioned, for instance, the system may be conceptually – and technically, actually, as explained in the following – split in a few layers, corresponding to the different scales at which it is conveniently modelled and designed; let us assume three as depicted in Fig. 3:

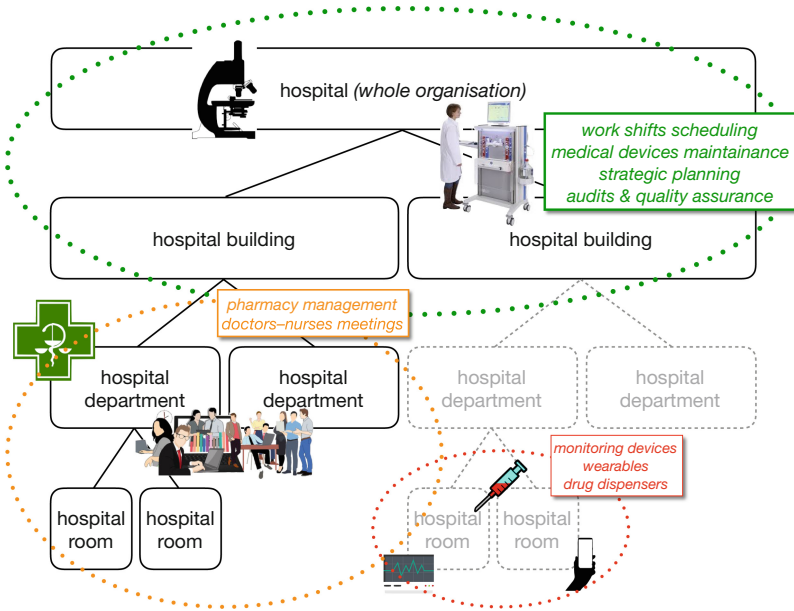
- the smaller scale is mostly concerned with local-only, critical, highly dynamic situations recognition and decision making (i.e. a single room where a patient may unexpectedly need the emergency unit)
- the medium scale is possibly the most difficult to define, since it is essentially meant to transition from the local perspective of the smaller one to the global-perspective of the larger one. Here, the most critical task is that of defining how information coming from the lower layer (the smaller scale) can be aggregated and presented to the upper layer (the larger scale), and how decision making executed on the higher layer should be translated in actionable commands for the lower one. For instance, coordination amongst doctors and nurses in the same department based on scheduled appointments and emergency events is likely to happen here
- the larger scale deals with global planning and monitoring, where collection of relevant aggregated information and synthesis of consequential activities happen on a medium to long-term horizon, and responsiveness is usually far less important than accuracy and completeness (of both information collection and decision making). This scale may range from an individual hospital building up to the whole hospital organisation as displaced in different geographical areas—but belonging to the same administration.

## 5.2 Middleware

Under a more pragmatic perspective, a crucial technical question is to understand the role of middleware in supporting the new means of coordinating distributed components, represented by conversations. In fact, although conversation essentially amounts to message-passing interaction, a mere message-oriented middleware (MOM) would fail addressing its peculiarities [6]. Conversations imply a shared knowledge among interacting components, which cooperatively build upon it a common interpretation of the world based on *logically sound* and related arguments, and cooperatively conceive and commit to a joint plan of actions. MOM is also weak in supporting interaction in a dynamic (i.e. open and mobile) world, where the identities and locations of components are not known in advance, as in the case of speaking objects (and of IoT in general).

Accordingly, the middleware should lean towards a different coordination model, capable of going beyond the rather primitive functionality of MOMs

in terms of direct interactions between components. Rather, it should support conversations at an higher level of abstraction, i.e. via an open and shared conversation space enabling conversation among components that do not necessarily have to know each other in advance: for instance, a tuple space. However, unlike traditional tuple space models, which contain unrelated pieces of data, the need to access data and metadata 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 [26], the best way to realize such shared conversation space is still subject of active research. As it is yet to be evaluated how corpora of commonsense knowledge could be integrated within the overall architecture to support conversations.



**Fig. 3.** Different scales of information collection, decision making, and coordination as seen in a large-scale Speaking Objects deployment. Smaller scales are associated with critical, highly dynamic situations, in which argumentation-based coordination may be employed to guarantee soundness and accountability of solutions, whereas larger scales with longer planning and monitoring, and slower but steady adaptation given by self-organising coordination may come handy to manage complexities.

### 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 Fig. 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, 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 conversing speaking objects.

In this perspective, more natural interfaces, such as voice commands or gestures, and techniques coming from natural language processing, speech recognition, and computer vision will become essential components of smart objects, 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 [21]. While the current state of the art is about interacting with a single device or hub (e.g., Amazon’ Echo and Google Home), 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.

Besides the need for effective means of human-machine interaction, as already discussed in Sect. 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

Nowadays, the world in which we are living is becoming more and more dominated by algorithms, that by now are daily exploited in a variety of decision-making processes. This novel scenario is typically referred to as an *algocracy* [7]. In such a framework, it is often the case that we act as passive subjects in situations that have been automatically planned and arranged for us by algorithms. This could become a crucial issue in the forthcoming years, when these systems will become a reality also on a large scale, for example in the context of smart cities, where the safety and well-being of citizens will largely depend on technology [41].

The scenario of speaking objects moves a step towards an open and *interpretable* network of smart devices, with which humans can naturally interact and converse, eventually understanding the choices and decisions of these agents, through argumentation and dialogue. These innovative elements provide a means through which it could be possible to control algocracy, by creating “grey-boxes”

whose behavior will be intelligible by an external observer that needs to inspect their way of acting.

## 5.5 Security

Distributed scenarios for IoT have been extensively studied in terms of security. Many challenges arise in a massive-scale scenario, including authentication, privacy preservation, data integrity, fault tolerance, trust, and governance [33]. The inherent nature of speaking and hearing objects is grounded on conversation. On the one hand, this makes the framework vulnerable to possible system intrusions and attacks, but at the same time it can represent a major advantage against malicious behavior, thanks to interpretable explanations given by speaking objects via argumentation. The research in the field of argumentation-based risk assessment [15] could be turned into automated argumentation-based security. At the same time, the correctness, validity, and strength of the posed arguments could be exploited to assess the reputation of speaking objects, and thus to enforce the concept of trust in the IoT setting.

## 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 for bringing in the lessons of massive multiagent systems. Within this new scenario, scalability will soon become a urgent need, which will require the integration of a number of technologies from different research areas. On the one hand, speaking objects will have to implement coordination through learning, reasoning, and especially argumentation, in order to show a behavior easily interpretable also for humans. On the other hand, such a large-scale scenario represents an ideal testbed for novel technologies in the field of distributed and pervasive computing, which will face challenges in the area of software engineering, security, and human-computer interaction.

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

## References

1. Agrawal, H., Leigh, S.-W., Maes, P.: L'evolved: autonomous and ubiquitous utilities as smart agents. In: ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 487–491. ACM, New York (2015)
2. Amgoud, L., Parsons, S.: Agent dialogues with conflicting preferences. In: Meyer, J.-J.C., Tambe, M. (eds.) ATAL 2001. LNCS (LNAI), vol. 2333, pp. 190–205. Springer, Heidelberg (2002). [https://doi.org/10.1007/3-540-45448-9\\_14](https://doi.org/10.1007/3-540-45448-9_14)
3. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. *Comput. Netw.* **54**(15), 2787–2805 (2010)



4. Bordini, R., et al.: A survey of programming languages and platforms for multi-agent systems. *Informatica (Ljubljana)* **30**(1), 33–44 (2006). Cited by 152
5. Conti, M., et al.: Looking ahead in pervasive computing: challenges and opportunities in the era of cyber-physical convergence. *Pervasive Mob. Comput.* **8**(1), 2–21 (2012)
6. Curry, E.: Message-oriented middleware. In: *Middleware for Communications*, pp. 1–28 (2004)
7. Danaher, J.: The threat of algocracy: reality, resistance and accommodation. *Philos. Technol.* **29**(3), 245–268 (2016)
8. Davis, E., Marcus, G.: Commonsense reasoning and commonsense knowledge in artificial intelligence. *Commun. ACM* **58**(9), 92–103 (2015)
9. Endler, M., Briot, J.-P., Silva e Silva, F., De Almeida, V.P., Haeusler, E.H.: 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)*, pp. 348–353. IEEE, San Diego, January 2017
10. Freitas, A., Bordini, R.H., Meneguzzi, F., Vieira, R.: Towards integrating ontologies in multi-agent programming platforms. In: *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, vol. 3, pp. 225–226. IEEE (2015)
11. Ganz, F., Puschmann, D., Barnaghi, P., Carrez, F.: A practical evaluation of information processing and abstraction techniques for the internet of things. *IEEE Internet Things J.* **2**(4), 340–354 (2015)
12. Garcez, A.S.D., Broda, K.B., Gabbay, D.M.: *Neural-Symbolic Learning Systems: Foundations and Applications*. Springer Science & Business Media, London (2012)
13. Getoor, L., Taskar, B.: *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*. The MIT Press, Cambridge (2007)
14. Gyrard, A., Serrano, M., Atemezing, G.A.: Semantic web methodologies, best practices and ontology engineering applied to internet of things. In: *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, pp. 412–417. IEEE (2015)
15. Ingolfo, S., Siena, A., Mylopoulos, J., Susi, A., Perini, A.: Arguing regulatory compliance of software requirements. *Data Knowl. Eng.* **87**, 279–296 (2013)
16. Islam, S.R., Kwak, D., Kabir, M.H., Hossain, M., Kwak, K.-S.: The internet of things for health care: a comprehensive survey. *IEEE Access* **3**, 678–708 (2015)
17. Jennings, N.R., Faratin, P., Lomuscio, A.R., Parsons, S., Wooldridge, M.J., Sierra, C.: Automated negotiation: prospects, methods and challenges. *Group Decis. Negot.* **10**(2), 199–215 (2001)
18. Jung, H., Tambe, M., Kulkarni, S.: Argumentation as distributed constraint satisfaction: applications and results. In: *Proceedings of the Fifth International Conference on Autonomous Agents, AGENTS 2001*, pp. 324–331. ACM, New York (2001)
19. Kephart, J.O., Chess, D.M.: The vision of autonomic computing. *IEEE Comput.* **36**(1), 41–50 (2003)
20. Kortuem, G., Kawsar, F., Sundramoorthy, V., Fitton, D.: Smart objects as building blocks for the internet of things. *IEEE Internet Comput.* **14**(1), 44–51 (2010)
21. Kranz, M., Holleis, P., Schmidt, A.: Embedded interaction: interacting with the internet of things. *IEEE Internet Comput.* **14**(2), 46–53 (2010)
22. Liggins II, M., Hall, D., Llinas, J.: *Handbook of Multisensor Data Fusion: Theory and Practice*. CRC Press, Boca Raton (2017)
23. Lippi, M., Mamei, M., Mariani, S., Zambonelli, F.: An argumentation-based perspective over the social IoT. *IEEE Internet Things J.*, 1 (2017)

24. Lippi, M., Mamei, M., Mariani, S., Zambonelli, F.: Coordinating distributed speaking objects. In: 37th IEEE International Conference on Distributed Computing Systems, ICDCS 2017, Atlanta, USA, 5–8 June 2017
25. Maarala, A.I., Su, X., Riekkki, J.: Semantic reasoning for context-aware internet of things applications. *IEEE Internet Things J.* **4**(2), 461–473 (2017)
26. Mariani, S.: Coordination of Complex Sociotechnical Systems - Self-organisation of Knowledge in MoK. *Artificial Intelligence: Foundations, Theory, and Algorithms*. Springer, Cham (2016). <https://doi.org/10.1007/978-3-319-47109-9>
27. Meneguzzi, F.R., Zorzo, A.F., da Costa Móra, M.: Propositional planning in BDI agents. In: Proceedings of the 2004 ACM Symposium on Applied Computing, SAC 2004, pp. 58–63. ACM, New York (2004)
28. Omicini, A., Viroli, M.: Coordination models and languages: from parallel computing to self-organisation. *Knowl. Eng. Rev.* **26**(1), 53–59 (2011)
29. Perera, C., Zaslavsky, A., Christen, P., Georgakopoulos, D.: Context aware computing for the internet of things: a survey. *Commun. Surv. Tutorials* **16**(1), 414–454 (2014)
30. Rahwan, I., Ramchurn, S.D., Jennings, N.R., Mccburney, P., Parsons, S., Sonenberg, L.: Argumentation-based negotiation. *Knowl. Eng. Rev.* **18**(4), 343–375 (2003)
31. Rao, A.S., Georgeff, M.P.: BDI agents: from theory to practice. In: Lesser, V.R., Gasser, L. (eds.) 1st International Conference on Multi Agent Systems (ICMAS 1995), pp. 312–319. The MIT Press, San Francisco, 12–14 June 1995
32. Razzaque, M.A., Milojevic-Jevric, M., Palade, A., Clarke, S.: Middleware for internet of things: a survey. *IEEE Internet Things J.* **3**(1), 70–95 (2016)
33. Roman, R., Zhou, J., Lopez, J.: On the features and challenges of security and privacy in distributed internet of things. *Comput. Netw.* **57**(10), 2266–2279 (2013)
34. Shi, W., Dustdar, S.: The promise of edge computing. *Computer* **49**(5), 78–81 (2016)
35. Tsai, C.-W., Lai, C.-F., Chiang, M.-C., Yang, L.T., et al.: Data mining for internet of things: a survey. *IEEE Commun. Surv. Tutorials* **16**(1), 77–97 (2014)
36. Walton, D., Krabbe, E.: *Commitment in Dialogue: Basic Concept of Interpersonal Reasoning*. State University of New York Press, Albany (1995)
37. Wooldridge, M.J.: *Reasoning About Rational Agents*. MIT press, Cambridge (2000)
38. Ye, J., Dobson, S., McKeever, S.: Situation identification techniques in pervasive computing: a review. *Pervasive Mob. Comput.* **8**(1), 36–66 (2012)
39. Yi, S., Li, C., Li, Q.: A survey of fog computing: concepts, applications and issues. In: Proceedings of the 2015 Workshop on Mobile Big Data, Mobidata 2015, pp. 37–42. ACM, New York (2015)
40. Zambonelli, F., Jennings, N.R., Wooldridge, M.: Developing multiagent systems: the Gaia methodology. *ACM Trans. Softw. Eng. Methodol.* **12**(3), 317–370 (2003)
41. Zambonelli, F., Salim, F., Loke, S.W., De Meuter, W., Kanhere, S.: Algorithmic governance in smart cities: the conundrum and the potential of pervasive computing solutions. *IEEE Technol. Soc. Mag.* **37**(2), 80–87 (2018)
42. Zambonelli, F., et al.: Developing pervasive multi-agent systems with nature-inspired coordination. *Pervasive Mob. Comput.* **17**(Part B), 236–252 (2015)
43. Zatelli, M.R., Hübner, J.F., Ricci, A., Bordini, R.H.: Conflicting goals in agent-oriented programming. In: Proceedings of the 6th International Workshop on Programming Based on Actors, Agents, and Decentralized Control, AGERE, pp. 21–30 (2016)