This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.

Emergence of proto-sentences in artificial communicating systems

Ryoko Uno , Davide Marocco , Stefano Nolfi and Takashi Ikegami

***Abstract*—This paper investigates the relationship between embodied interaction and symbolic communication. We report about an experiment in which simulated autonomous robotic agents, whose control systems were evolved through an artificial evolutionary process, use abstract communication signals to coordinate their behavior in a context independent way. This use of signals includes some fundamental aspects of sentences in natural languages which are discussed by using the concept of joint attention in relation to the grammatical structure of sentences.**

***Index Terms*— artificial life, communication with and without language, sentences, joint attention.**

I. Introduction

WHAT are the precursory variations in preverbal communication before language system emerges? Chomsky could argue that no such precursory system exists. However, recent developments in cross-disciplinary studies have revealed that there are many examples showing the embodied groundings for language communication. Kita, who studied Japanese mimetics [1], [2] argues that the meaning of mimetics is primarily represented in an *affect-imagistic* dimension where “language has direct contact with sensory motor and affective information” [1] and “vivid imagery of perceptual and physiological experiences” [2]. Glenberg [3] shows how actions execution interferes with understanding the meaning of sentences. For example, “Open the door!” facilitates the act of pulling and, conversely, “Close the drawer” facilitates the act of pushing.

The relationship between language and embodiment has been intensively sought after since the discovery of mirror neurons [4]. The so-called mirror neural system (MNS) is defined as a neural subsystem responsible for matching the self-action with the same intentional action performed by someone else. Rizzolati and Arbib [5] have argued that a ventral premotor cortex called F5 and AIP found in Makaque monkeys might have played a key role for language evolution. In addition to reporting functional similarities between F5 and the human Broca area, they hypothesize a unified scenario where the MNS is the bootstrapping factor at the root of any communication system in animals and humans, which they call an observation/execution matching system.

If language is embodied in nature, the function cannot be limited to information transmission, that is to merely use language as a tool. As in the case of bodily communication, language communication is used just to communicate and share the intentionality, not explicit information. For example, when someone says “The moon is beautiful tonight” looking up to the sky, it is not only meant to notify the beauty of the moon as information, but also to confirm that they are sharing the same experience. Since the moon is in front of the two, the information is redundant. It is a sentence used to share the context, and the mental state. And even without a physical element that can be shared by two people, such as the moon, language can be used to induce shared intentionality context.

In this paper, with an artificial life approach, we aimed to investigate this latter usage of language. When agents are evolved with uncertain information, their communicative interaction shows the following two characteristics: (1) signals are used to share the situation and (2) signals are detached from the context.

In the following sections, after a brief explanation of the framework used in the present work, a new evolutionary model and its results will be reported. Finally, in the discussion section, some arguments will be presented about similarities that can be found between the nature of linguistic sentences and some characteristics observed in the experiment reported.

II. Artificial Life Approach

In the present work, we adopt an artificial life approach to explore the relationship between embodiment and language. Indeed, artificial life studies provide a test bed for exploring how symbols and grammars emerge in minimal interacting systems through computer simulation. For the last 10-15 years, artificial life studies have contributed greatly to this direction, and the origin and evolution of language has become a target of many scientific study (see e.g. [6], [7], [8], [9], [10], [11], [12], etc.). For example, Steels and Kaplan [13] have developed a platform for studying the interaction between two artificial

Manuscript received December 17, 2010. This work was supported in part by the European Commission FP7 Project ITALK (ICT-214668) within the Cognitive Systems and Robotics unit (FP7 ICT Challenge 2)

R.Uno is with Institute of Technology, Tokyo University of Agriculture and Technology (phone: 81-42-388-7593; e-mail: ryokouno@cc.tuat.ac.jp).

D. Marocco is with the Centre for Robotics and Neural Systems, University of Plymouth, Plymouth, United Kingdom (e-mail:davide.marocco@plymouth.ac.uk).

S. Nolfi is with the Institute of Cognitve Science and Technologies, National Research Council, Rome, Italy (stefano.nolfi@istc.cnr.it).

T. Ikegami is with the Department of General Systems Sciences, The Graduate school of Arts and Sciences, University of Tokyo (ikeg@sacral.c.u-tokyo.ac.jp).

agents acting as speaker or hearer. In this approach a population of robots develop a shared vocabulary and a corresponding ontology (including eventually grammatical words) while playing language games (i.e. ritualized social interactions that follow a specific script). For example, by performing a “naming” game, which consists in trying to bring the attention of another robot on a given object located in the environment through linguistic interactions, robots develop a list of symbolic concepts (associated to words and to perceptual features) with which they can refer to the objects. This approach thus is based on the idea that language is a process in which the speaker maps sub-symbolic sensory experiences into symbols and then symbols into words (or eventually sentences) and in which the hearer performs the inverse mapping. The functional use of language is then performed by a separate module of the robots’ control system that takes care of how the symbolic information conveyed through language should influence the robots’ behavior.

Di Paolo [14], as well as, Marocco and Nolfi [15], [16] investigated how a communication system can emerge in groups of robots provided with neural controllers that are evolved for the ability to solve problems that require cooperation. In particular, Marocco and Nolfi studied an experimental scenario where, four wheeled robots situated in an arena containing two circular target areas painted in black, are evolved for the ability to reach and remain in the black areas by equally subdividing between the two areas. The robots can interact between themselves by detecting the presence of an obstacle constituted by another robot through their infrared sensors, and by producing and detecting acoustic signals. The analysis of the obtained results shows how evolving robots develop a structured communication system and use such a system in order to cooperate and successfully solve their task.

The analysis of the evolutionary process [15] indicates that during a first phase robots develop an ability to explore the environment. Consequently, this ability enables them to eventually reach the target areas. At this stage the robots are not able to stop their exploratory behavior and remain within the areas. During this initial phase the robots produce signals and react to detected signals in a way that do not increases their performances.

During a second phase, robots develop an ability to exploit their ground sensors (that detect the color of the floor) to remain on target areas and to vary the signal produced inside and outside target areas. This creates the condition for the development of a first form of communication, in which the robot located inside the target area produces a signal that provides information about the location of the area itself, while the robots nearby reacts to the signal by turning and heading toward the direction of the target area. This leads to a form of mono-directional communication in which the robot located inside the target area acts as a speaker (i.e. it produces the signal that influence other robots but is not influenced by other robots’ signals) while the robot located outside target area acts as a hearer (i.e. it does not produce signals and reacts to detected signal by modifying its behavior).

During a third phase, the robots develop an additional form of communication that allows them to establish how many robots are located inside a target area and to differentiate the type of signal produced on the basis of how crowded the current area is. This is achieved through a bi-directional communication form in which the robots located into a target area concurrently act as speaker and hearer (i.e. they produce signals that modify the behavior of the other robot and at the same time are influenced by the signal produced by the other robot). This bi-directional communication form allows (a) robots located outside target areas to avoid moving toward crowded areas and (b) robots located in target areas containing more than two robots, to abandon their area in order to ensure that the group of robots will equally divide between the two areas.

Other studies (e.g. [17], [18]) have investigated how communication can emerge also in agents that can interact through their motor behavior only (i.e. that do not have the possibility to exploit a dedicated communication channel).

One study that is particularly interesting from the point of view of the transition between non-verbal and verbal communication is the study of IIzuka and Ikegami’s [19], [20] in which evolving agents coordinate in order to realize a turn-taking behavior by exploiting subtle variations of their actions rather than explicit cues. In this work, turn-taking is formulated as a chasing spatial game where an agent has to follow the other agent in a two dimensional arena. The follower/leader role played by the agents depends on their relative position in space. The main purpose of the simulation was the development of agents that can peacefully switch between a follower and a leader role within a given time frame. Moreover, the switching behavior should be self-emerging without having explicit cues or signs. The interest of this work is in the basic mechanisms of mutual interaction between two agents which allows them to autonomously switch the role. The dynamic process seen in the formulation of turn-taking and in joint attention mechanisms is also studied by Ito and Tani [21] on a human-robot interaction setting, where the authors analyzed the sensory-motor patterns emerged during these interactions. They observed a spontaneous emergence of synchronized and unsynchronized phase movements between the human and the robot, that are linked to essential psychological mechanisms of joint attention. This view was then theorized and critically explained in [22] by Tani.

In this paper we extended the experiments performed by Marocco and Nolfi [15], [16] to investigate the relationship between non-verbal and verbal interactions and the relevance of such a relationship in the context of language evolution. Notably, the importance of non-verbal and verbal interactions has inspired parallel works on robotics and artificial agents, whose attempt to exploit those interactions to bootstrap “intelligent” coordinated behaviours in a human-robot interaction context [23].

In the experiment reported here, in addition to the analysis performed on robots’ coordinated behavior, we attempt to reveal the linkage between synchronized behavior and symbolic aspects of language, that is, the detachment of signals

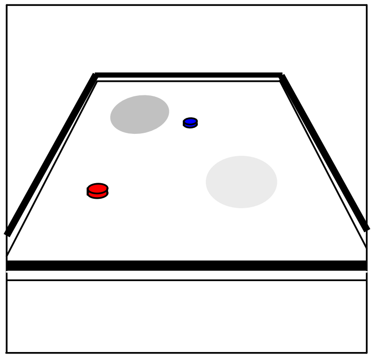


Fig. 1. The simulated environment with the two robots (small circles) and the two target areas with different gray shades.

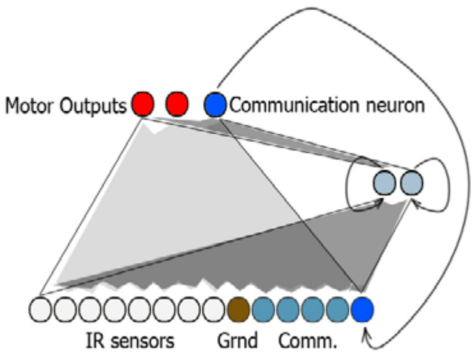


Fig. 2. The architecture of the neural controller of the robots. The grey areas indicate the connections between blocks of neurons. Arrows indicate the recurrent connections of the communication output and of internal units.

from the context in which they are produced. As we have reviewed in this section, symbolic language features and coordinated behavior has been studied separately in the context of A-life. Thus, a major contribution of this work is to link these two topics in order to modeling the relationship between verbal and non-verbal interactions and discover the inner link between behavior coordination and symbolic communication.

III. A Simulation Model of Evolution of Communication

To investigate the hidden link between behavioral and signal coordination, a new experiment based on a modification of [16] set up was carried out. In modifying Marocco and Nolfi's set up, three major changes were made: (1) during the training phase the target area can be any shade of gray and can change the color at every trial; (2) only two agents instead of four are present in the environment; (3) robots can hear each other's signals all over the environment, while in Marocco and Nolfi's original model the signaling communication was limited to the local neighborhoods.

A. *Experimental Scenario*

The experimental scenario involves two wheeled robots which are placed in a 150x150cm arena surrounded by walls (fig. 1). The arena has a white ground and contains two target areas painted in different grey tonalities. The robots are evolved for the ability to find and remain in the same target area (i.e. in one of the two areas).

Robots' neural controller consists of a neural network with 14 sensory neurons, 2 internal neurons with recurrent connections, and 3 motor neurons (fig. 2). The sensory neurons encode the activation states of the corresponding 8 infrared sensors (which provide information about obstacles, i.e. walls or the other robot located nearby up to a distance of 5 cm), 1 ground sensor (which provides information about the colour of the floor), 4 communication sensors (which encode the value of signal detected and direction of the source of the signal within four corresponding intervals: frontal [315°-44°], rear [135°-224°], left [225°-314°], right [45°-134°]), and 1 sensory neuron (which encodes the signal produced by the robot itself at time t-1). The motor neurons encode the desired speed of the two wheels and the signal produced by the robot. Signals consist of floating point values in the range [0.0, 1.0] and can be detected from any distance within the limit of the arena.

The output of motor neurons is computed according to the logistic function (2), the output of sensory and internal neurons is computed according to function (3) and (4), respectively (for more details on these activation functions and on the relation with other related neural models see Nolfi, 2002).

$$A_j = t_j + \sum_i w_{ij} O_i \tag{1}$$

$$O_j = \frac{1}{1+e^{-Aj}} \tag{2}$$

$$O_j = O_j^{(t-1)} \tau_j + I_j (1-\tau_j) \tag{3}$$

$$O_j = O_j^{(t-1)} \tau_j + \left(1+e^{-A_j}\right)^{-1} (1-\tau_j) \tag{4}$$

With A_j being the activity of the j_{th} neuron, t_j being the bias of the j_{th} neuron, w_{ij} the weight of the incoming connections from the i_{th} to the j_{th} neuron, O_i the output of the i_{th} neuron, $O_{j(t-1)}$ being the output of the j_{th} neuron at the previous time step, τ_j the time constant of the j_{th} neuron, and I_j the activity of the j_{th} sensors.

The free parameters of the robots' neural controllers have been evolved through a genetic algorithm. Each group of two robots was tested for 20 trials, lasting 120 seconds each (i.e. 1200 cycles of 100 ms). At the beginning of each trial the colour of the two target area was randomly assigned in the range [0.0, 1.0] (where 0.0 corresponds to white, i.e. the same colour of the rest of the ground, 1.0 corresponds to black, and intermediate levels correspond to different grey colors) and the robots are placed in a randomly selected position and orientation outside the target areas. The fitness of the group consists of the sum of 0.1 scores for each robot located in a target area alone and of 0.5 scores for each robot located in a target area together with the other robot. The total fitness of a group is computed by summing the fitness gathered by the two robots in each cycle.

The initial population consisted of 100 randomly generated genotypes that encoded the connection weights, the biases, and

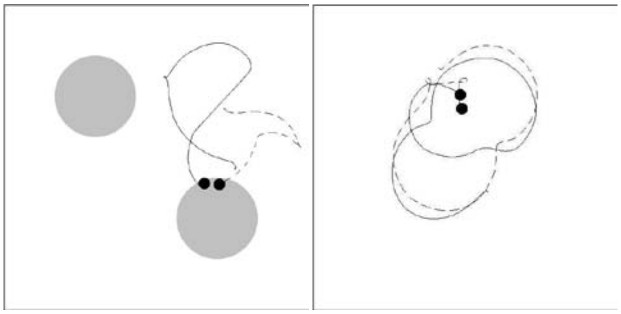


Fig. 3. The behavior observed in two different environmental conditions: left) the environment contains two black target areas and (right) the environment does not contain any target area. That is, target areas are white, therefore indistinguishable from the rest of the ground. Solid and dashed lines represent the trajectories of the two robots (small black circles).

the time constants of 100 corresponding neural controllers. Each parameter was encoded with 8 bits and normalized in the range [–5.0, +5.0], in the case of connection weights and biases, and in the range [0.0, 1.0], in the case of time constants. Each genotype was translated into two identical neural controllers that were embodied in the two corresponding robots, i.e. teams were homogeneous and consisted of two identical robots. For a discussion about this point and alternative selection schemas see [17] and [18]. The 20 best genotypes of each generation were allowed to reproduce by generating five copies each, with 2% of their bits replaced with a new randomly selected value. The evolutionary process lasted 300 generations (i.e. the process of testing, selecting and reproducing robots is iterated 300 times).

The experiment was replicated 10 times with different initial population randomly generated.

It is interesting to note that, from a perceptual point of view, the way in which the experimental scenario is built implies that sensory values greater than 0.0 indicate that the robot is located in a target area while values equal to 0.0 indicate that the robot is located outside target areas. Since the color of the target area can be any value between 0 and 1, the colour of the target area by itself does not provide a reliable indication whether the two robots are both located in the same target area or they are located in a target area at all. Indeed, as we will see, robots need to develop a sophisticated communication ability to differentiate these two conditions.

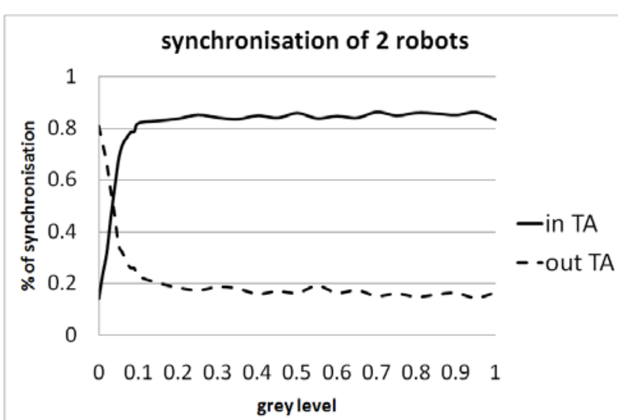


Fig. 4. The amount of synchronization events inside the target area (solid line) and outside (dashed line). On the *x* axis the different grey levels of the target area, from white (0) to black (1).

B. Simulation Outcomes

The observation of the behaviors evolved in the 10 replications shows that in all the cases robots are able to perform correctly the task, although the communication and behavioral strategies differ to a certain extent for all the replication.

The large majority of the replications show a good ability of the robots in (a) exploring the environment and locating target areas; (b) a proficient use of acoustic communication signals to locate the position of the robot already inside the target area and (c) an ability to use their bodily motion to identify whether they are located on the same area or not.

Besides those general results, we are particularly interested in a specific strategy observed in only one replication. It is worth to mention that the fitness performance of this replication, although not the best, is comparable with the overall performances of the 10 replications: *Average best teams* for 10 replications: 615.78±52.4(599.91); *average mean population* for 10 replications: 333.52±75.05(291.63) - within brackets the analyzed replication.

What has been observed is that, in this particular case, sometimes the two robots synchronize and coordinate their behavior not only inside the target area, but also outside. This is something that cannot be predicted from the fitness function and it is the result of the uncertainty in the color of the ground. Figure 3 shows the behavior observed in two different environmental conditions: (a) The environment contains two grey target areas and (b) the environment does not contain any target area. Thus, target areas are white, therefore indistinguishable from the rest of the ground. Figure 3 shows the trajectories of the two robots in the two conditions.

To confirm this observation and measure to what extent the coordination outside target areas was achieved under variable environmental conditions, a test was run in which the synchronization events (that is, the time in which the two robots mutually engaged in a coordinated behavior) were recorded by testing the robots under different environmental conditions, i.e. the color of the target area was systematically changed from 0 (white) to 1 (black). The graph in figure 4 shows that the majority of the synchronization events outside target areas are observed only when the color of the area is white or very close to it, and the number of synchronizations inside the target areas definitively outperforms the synchronizations outside as soon as the color of the area is effectively distinguishable from the ground.

In order to better understand the different roles played by different sensors in producing behavioral synchronization, an additional test was performed. In the test, the amount of synchronization events was recorded by systematically testing all the possible sensors' combinations. The environment was arranged with only one black target area and testing trial lasted for 5000 timesteps (or cycles). Only the synchronization events

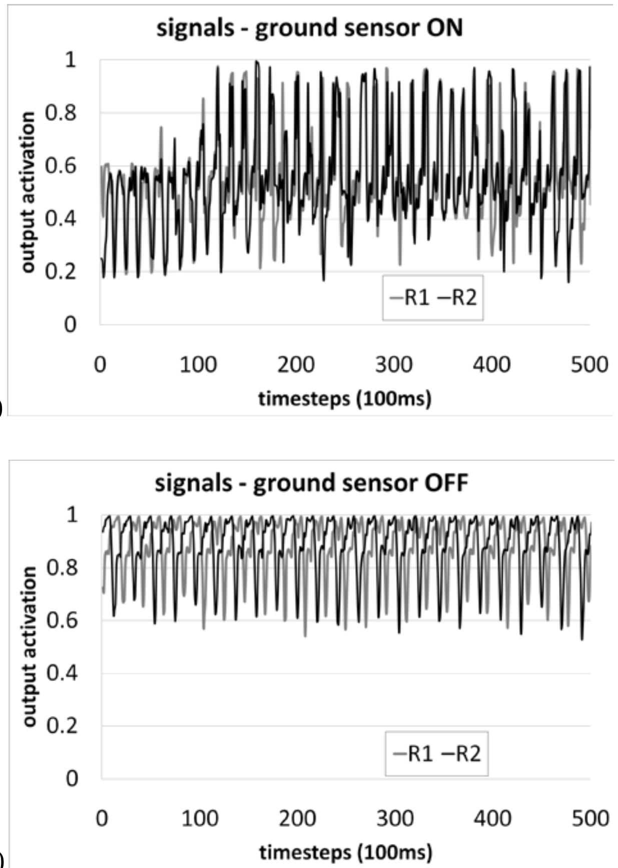


Fig. 5. Signals emitted by two interacting robots (black and grey lines) over time in two different conditions: (a) with the ground sensor activated inside a visible target area and (b) with the ground sensor not activated (outside a target area, or over a not visible target area).

that lasted up to the end of the trial was recorded. 1000 trials for each condition were performed.

Table I shows the results of such test, in which we can clearly observe that, as it was expected, the only redundant sensor is the ground sensor (***Ground***). Indeed, with and without the ground sensors (which is comparable to visible or not-visible target areas in the arena) the two robots can establish synchronization inside and outside the areas. On the other hand, without the communication sensor (***Comm***), that prevents the robots to detect signals emitted by the other one, synchronization is never observed. Therefore, the communication channel is crucial for the observed synchrony of behaviors, both inside and outside the target areas. Interestingly, infrared sensors (***IR***) appear to be also very important for bootstrapping synchronization, since without infrared sensors synchronization is never observed. In this specific condition, a visual observation of behaviors revealed that in cases in which ***Comm*** was on, robots started to synchronize thanks to the communication channel, but after a while they would crash onto each other (data not shown).

Since the communication channel plays the most important role in allowing behavioral synchronization, in order to understand how communication signals were exchanged between robots in the two conditions, i.e. with the ground sensor active or not active, an additional test was ran. In this test two robots were interacting in an environment with only one target area, that was white or black, and the communication signals emitted by the two robots were recorded while they were interacting freely. Figure 5 shows the graphs of the two interacting signals in two different conditions: (a) with the ground sensor activated (inside a visible target area) and (b) with the ground sensor not activated (outside a target area, or within a not visible target area). It is interesting to note that behavioral synchronization is achieved through different communication strategies. Pattern of signals exchange are different in the two conditions and differ both in terms of amplitude and phase of the oscillation. In (a) signals are in phase and the amplitude spans in the range [0.2-1]. On the contrary, in (b), the amplitude is confined in the smaller range [0.6-1] and the two signals are in anti-phase.

TABLE I
BEHAVIORAL SYNCHRONIZATION EVENTS AND SENSOR DEVICES

IR	*Ground*	*Comm*	*% sync*	*In*	*Out*
off	off	off	0	0	0
off	off	ON	0	0	0
off	ON	Off	0	0	0
off	ON	ON	0	0	0
ON	off	off	0.4	0	4
ON	off	ON	95.2	143	809
ON	ON	Off	0.4	0	4
ON	ON	ON	100	833	167

Summary of tests done with different sensors configurations. *IR* - infrared sensors, *Ground* – ground sensor, *Comm* – Communication output; *% sync* – percentage of synchronization events recorded over 1000 trials. *In* – events recorded inside the target area, *Out* – events recorded outside the target area.

IV. OBSERVATIONS

A first comment about the results obtained and the analysis performed in this new setup is that infrared sensors and communication channel synchrony is more unstable than in the case of a reliable ground sensing, as in the case of Marocco and Nolfi's experiment. Since robots cannot trust the information provided by the ground sensor, they evolved an interaction strategy that allows them to be entrained in an imaginary synchrony even in the absence of a target area. This ability of creating imaginary context and sharing behavioral patterns only based on embodied interactions and abstract communication signal is, in our point view, an essential function of language. Indeed, with respect to the original Marocco and Nolfi's experiment, the new setup allows to observe different "strengths" in the symbols that mediate the collective behavior. This happens because the signal for the target area is uncertain and the robots try to use other sensory channels to accomplish their collective task. In particular, infrared sensors change their meanings. Since there are only two agents in the new experiment, when infrared sensors are activated it means either that the other agent or a wall is in the nearer neighborhood of a robot. In Marocco and Nolfi setup, robots were not able to use the infrared sensors to detect other robot in the target area. They were only able to sense the presence of another robot as an obstacle in the environment and then avoid it. In this new setup, given that only two robots are present in the environment, detecting the presence of another robot by means of the infrared sensors becomes possible. In particular, in order to discriminate between a wall and the other robots, synchronization becomes a very useful strategy. In this respect we also observed that inputs from infrared sensors generally override the input from the ground sensor, i.e., only infrared sensors are capable to trigger synchronization (see results in Table I). In this case, synchronization is exclusively used to find out whether the robot is interacting with another robot or not. Indeed, it seems that communication signals and infrared sensors are used in cooperation to build up a ground of interaction between the two robots.

The use of synchronization to discriminate between walls (i.e. passive obstacles) and robots was tested by actively modifying the infrared sensors activations of the robots while they were synchronized. Robots were tested for 2000 trials of 1000 timesteps. Communication signals remained untouched. In this testing condition, if the sensory appearance of one of the two robots is suddenly substituted with a wall (1000 trials) or a fixed round obstacle (1000 trials), the infrared sensors pattern suddenly changes and the interacting robots, the non-modified one, after few timesteps breaks up the synchronization and starts again the exploratory behavior (this happened the 100% of the times out of 2000 trial), despite the fact that the communication signals were not actively affected by the change. In fact, the infrared sensors modification induces an alteration on the communication interaction and, in turn, the synchronization behavior becomes unstable and brakes up. This effect is observed regardless the ground sensor activation.

If the synchronization inside a target area can be seen as a direct outcome of the fitness (robots must stay together in the area), synchronization events outside the target areas are only indirectly related to the fitness and a relatively rare outcome in the whole experiment (as we said, only 1 out of 10 replications). In this regard we can remark two interesting points: (1) synchronization of signals can be seen as a mean to induce coordinated behaviors between the two robots. This behavior coordination is known to facilitate the coordination of inner state, thus inducing intersubjectivity [24]; and (2) some particular sentences in language are very important because they can induce intersubjectivity without depending on specific contexts. In the following section these points will be discussed in detail. Taking those points into consideration, we can speculate that the present simulation shows some interesting characteristics of language that can be identified thanks to the embodied communication between robots. In the light of this hypothesis, we will explain the obtained results focusing on the following three points:

1) The relationship between the color of the ground and synchronization;
2) Synchronization outside the target area;
3) Signals used to synchronize outside and inside the target area.

From the results we can infer two different types of synchronization between the robots. In a first type, the synchronization of robots is obtained through the cooperation of all sensory channels inside the target area. We call this type of interaction: "synchronization inside a target area". In a second type, a synchronization of infrared sensors and communication signal is observed, while the ground sensor is not involved. This happens outside the target area and we call this interaction "synchronization outside the target area".

In the case of "synchronization inside a target area", at the beginning of a trial, both robots explore the environment looking for the target area. Then, one of them gets into the target area by chance while continuously emitting signals which the other robot can hear. At this point signals are not correlated, but when the one outside the target area gets inside (guided by the signal of other robot) they immediately start to synchronize their signals (figure 5a, 0-100 timesteps). They keep moving toward each other and after a while they get close enough to mutually activate the infrared sensors. At this point, the two communication signals show a continuous synchronization with an increment in amplitude (figure 5a, 100-500 timesteps).

On the other hand, in the case of "synchronization without a target area", the trial starts as usual with an exploration of the environment and then, by chance, the robots find each other outside a target area (or over a non-visible area). Once the infrared sensors of both robots are turned on, their communication signals start to synchronize. Namely, the triggering of the synchronization is always given by the communication channel, but only after the mutual activation of the infrared sensors, the communication signal can synchronizes (figure 5b). Therefore, the potential cue for establishing coordinated behavior and synchronization using communication signals are, in the order: ground sensor < infrared sensor < communicative channel.

V. Discussion and Conclusion

Based on the observation and the analysis performed on synchronization effects in the previous section, in the following sections those aspects will be analyzed from the point of view of communication and language studies.

A. Synchronization, behavior coordination and joint attention

In the experiment by Marocco and Nolfi [16] the usage of synchronized signals were observed only inside a target area. This type of communication is called by the authors *bi-directional communication*. On the other hand, the signals used by robots to indicate that there is a target area to other robots is called *mono-directional communication*. In the case of mono-directional communication, the information (i.e. the location of the target area) is transmitted to the robots outside the target area through the signal emitted by the robot inside the target area. The robot which listens to these signals can easily reach and get into the target area. Therefore, the authors speculate that signals emitted by the speaker have actively changed the behavior of the hearer in a mono-directional way. Information transmission is central to this communication. In contrast, the function of the bi-directional signals seems to be different. Bi-directional signals do not induce a new behavior of another robot but seems to be used to maintain the coordinated behavior of the two robots inside the target area. It is worth to note that both communication styles are equally relevant to the task achievement and both are functional and complementary to the realization of the observed final behavior.

The latter case in Marocco and Nolfi's simulation, in particular, is very similar to the "synchronization inside the target area" in the new setup. It should be noted that, in the case of bi-directional communication, the location of the target area is not informative, since both of them are already inside the target area. This general aspect of bi-directional communication is even more emphasized in the "synchronization outside the target area" in the new experiment. As there is no target area, robots signaling behavior becomes even less informative compared to the synchronization inside the target area, and, as we pointed out, robots rely much more on infrared sensors (IR), which is directly related to behavior coordination.

From psychological researches, we know that coordinating behavior in humans is deeply related to coordinating inner states: in other words, establishing "intersubjectivity" [24]. For establishing intersubjectivity it is supposed that there is a form of proto-linguistic communication called joint attention, which can be defined a coordinated preverbal behavior among two or more persons mediated by an object [25]. A simple example of joint attention is a children's pointing behavior under the attention of the mother. It is a process of sharing one's experience of observing an object or events with others by following pointing gestures or eye gazing that induces joint attention.

Following Bates' argument [26], two types of joint attention have been distinguished. In the first type, joint attention is used as a tool to achieve a goal (e.g. establishing joint attention to let your child pick up a toy). In the second type, joint attention itself is taken as a goal. For example, two people looking at the same sunset can establish this type of joint attention without requiring further achievements. The former is called "instrumental joint attention" and the latter "participatory joint attention" ([27], [19]). It is interesting to note that participatory joint attention, according to various authors, can only be achieved by normal human beings, and autistic children, for example, do not engage in participatory joint attention [25], [28], [29].

Two kinds of communication styles pointed out in Marocco & Nolfi [16] and in the experiment presented here, can have a correspondence with the two types of joint attentions: Mono-directional communication can be seen as a form of instrumental joint attention and bi-directional communication a form of participatory joint attention. Then, what is the difference between the two types of bi-directional communication (i.e. synchronizations inside and outside the target area) observed in the simulation?

B. *"Imaginary" joint attention with language*

The importance of joint attention in acquiring language is pointed out by Tomasello [25], [29] and others. However, for the purpose of the present work, it will be particularly investigated how the grammar of a mature language and joint attention are related.

A holophrase (a word whose function is like a sentence) is a kind of pointing, using a word instead of a finger or eye movement. A noun, for example, can be used to direct attention to a particular object [30]. In addition, since a person usually "presents" a word to another person, uttering a holophrase to someone is something similar to establishing joint attention through words. If someone says "Water!" it likely means that the person needs some water. This can establish instrumental joint attention between the speaker and the hearer. On the other hand, if someone says "Sun!" it probably means that the person is internally attracted by the sun (we are aware of the central role played by the context in those cases. Examples are intentionally over-simplified for the sake of clarity). If there is a hearer, this type of holophrase tends to establish participatory joint attention, rather than instrumental joint attention. On the same vein, it is also possible to share intersubjective states using a complete sentence whose meaning is redundant. For example, declarative sentences such as "Today is Sunday" or "Snow is white" can be used for this purpose.

Despite the fact that both holophrases and declarative sentences can be used to establish intersubjectivity, two main differences can be identified. Pointing using fingers or eyes, as well as holophrases, can be used only when the object that is pointed at or whose name is expressed, is present to the speaker. In this case, we can consider the pointing to be dependent on the ground, as well as an holophrase is dependent on the ground of speech [30]. On the contrary, complete sentences are free from the ground of speech, because what the speaker is going to present can be totally based on a knowledge that only belongs to the speaker. Because sentences that depend on grammar can convey complex relations, establishing intersubjectivity independently of the current situation becomes possible.

Furthermore, not only the dependence on the context differs, but there is a difference in the motivation by which intersubjectivity is achieved. When a holophrase is used, the relationship between the speaker and the object triggers the intersubjective state and the speaker is attracted and moved by the object itself [30]. On the contrary, sentences such as "Today is Sunday" or "Snow is white" are ground-independent in achieving shared intersubjectivity. The speaker attempts to draw the attention of the hearer to something other than themselves to maintain the communication. What it is important, is that the focus of shared attention does not have to be physically present. It can simply be a relationship which can be expressed through language. Thus, establishing intersubjectivity with declarative sentence can be called "imaginary participatory joint attention". Thanks to language, the "object" to which people pay attention can be imaginary and, interestingly, such an "imaginary object" is not what motivates joint attention. This kind of joint attention, which is purely induced by the existence of others, can be achieved with complete sentences but not with holophrases, because the detachment from the physical context is necessary.

Going back to our simulation, we can speculate that "synchronization without the target area" corresponds to imaginary participatory joint attention: First of all, there is no target area involved, but only through the mutual exchange of "abstract" signals, robots coordinate their behaviors "as if" there is a target area. Similarly, in the case of sentences, while there is no object to point at in the real world, people can achieve joint attention by talking about something that can be detached from the physical context. In addition, behavior coordination is achieved by robots, by strongly relying on infrared sensors, which means that what is central to the communication is the existence of each other. To coordinate the relationship between the two, imaginary participatory joint attention is induced. "Synchronization without target area" is free from the context and purely communicative, hence, it might indicate the emergence of proto-sentential signals.

In conclusion, what has been observed in the reported work is just the beginning of the grammar, that we believe has the potentiality to illuminate the relationship between non-verbal and verbal communication.

References

[1] S. Kita, "Two-dimensional semantic analysis of Japanese mimetics," *Linguistics*, vol. 35, pp. 379-415. 1997.

[2] S. Kita, "Semantic schism and interpretive integration in Japanese sentences with a mimetic: a reply to Tsujimura," *Linguistics*, vol. 39, pp. 419-436. 2001.

[3] A. M. Glenberg, and M. P. Kaschak, "Grounding language in action," *Psychonomic Bulletin and Review*, vol. 9, pp.558-565. 2002.

[4] G. Rizzolatti, L. Fadiga, V. Gallese and L. Fogassi, "Premotor cortex and the recognition of motor actions," *Cognitive Brain Research*, vol. 3, pp. 131-141. 1996.

[5] G. Rizzolatti, and M. A. Arbib, "Language within our grasp," *Trends in Neurosciences*, vol. 21-5, pp. 188-194. 1998.

[6] L. Steels, "The emergence and evolution of linguistic structure: from lexical to grammatical communication systems," *Connection Science*, vol. 17-3, 4, pp.213-230. 2005.

[7] S. Kirby, "Natural language from artificial life," *Artificial Life*, vol. 8-2, 185-215. 2002.

[8] A. Cangelosi and S. Harnad, "The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories," *Evolution of Communication*, vol. 4-1, pp. 117-142. 2000.

[9] P. Vogt. "Perceptual grounding in robot," in *Proceedings of the 6th European Workshop on Learning Robot*, A. Birk and J. Demiris, Eds. Springer-Verlag, 1998. 126-141.

[10] T. Hashimoto and T. Ikegami, "Emergence of net-grammar in communicating agents," *BioSystems*, vol. 38, pp. 1-14. 1996.

[11] Y. Sugita and J. Tani, "Learning semantic combinatoriality from the interaction between linguistic and behavioral processes," *Adaptive Behavior*, vol.13-1, pp. 133-52. 2005.

[12] K. Sasahara and T. Ikegami, "Evolution of birdsong syntax by interjection communication," *Artificial Life*, vol. 13, pp. 1-19. 2007.

[13] L. Steels and F. Kaplan, "AIBO's first words: The social learning of language and meaning," *Evolution of Communication*, vol. 4-1, pp. 3-32, 2001.

[14] E. A. Di Paolo, "Behavioural coordination in acoustically coupled agents," in *Proceedings of 8th International Conference on Artificial Neural Networks*, L. Nicklasson, M. Boden, and T. Ziemke, Eds, London; Springer, pp. 1097-1102. 1998.

[15] D. Marocco, and S. Nolfi, "Origins of communication in evolving robots," in *Proceedings of the 8th International Conference on Simulation of Adaptive Behavior*, Vol. 4095, S. Nolfi, G. Baldasarre, R. Calabretta, J.C.T. Hallam, D. Marocco, J. A. Meyer, O. Miqlino, and D. Parisi Eds. Springer-Verlag, 2006, pp 789--803.

[16] D. Marocco and S. Nolfi, "Emergence of communication in embodied agents evolved for the ability to solve a collective navigation problem," *Connection Science*, vol. 19, pp.53-74, 2007.

[17] G. Baldassarre, S. Nolfi, and D. Parisi, "Evolving mobile robots able to display collective behavior," *Artificial Life*, vol. 9, pp. 255-267, 2003.

[18] M. Quinn, "Evolving communication without dedicated communication channels," in *Proceedings of the 6th European Conference on Advances in Artificial Life*, J. Kelemen and P. Sosík, Eds. Prague: Springer, 2001, pp. 357-366.

[19] H. Iizuka and T. Ikegami, "Adaptability and diversity in simulated turn-taking behavior," *Artificial Life*, vol.10, pp. 361-378. 2004.

[20] T. Ikegami and H. Iizuka, "Turn-taking interaction as a cooperative and co-creative process," *Infant Behavior and Development*, vol. 30, pp. 278-288. 2007.

[21] M. Ito and J. Tani. "On-line imitative interaction with a humanoid robot using a dynamic neural network model of a mirror system," *Adaptive Behavior*, vol.12-2, pp. 93-115. 2004.

[22] J. Tani "Autonomy of Self at criticality: The perspective from synthetic neuro-robotics," *Adaptive Behavior*, 17-5, pp. 421-443. 2009.

[23] E. Bicho, L. Louro, and W. Erlhagen, "Integrating verbal and nonverbal communication in a dynamic neural field architecture for human-robot interaction," *Frontiers in Neurorobotics*, vol. 4-5. 2010.

[24] L. Murray and C. Trevarthen, "Emotional regulations of interactions between two-month olds and their mothers," in *Social Perception in Infants,* T. M. Field and N. Fox, Eds. Norwood: Ablex, 1993, pp. 177-197.

[25] M. Tomasello, *The Cultural Origins of Human Cognition*, Cambridge: Harvard University Press, 1999.

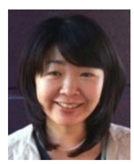
[26] E. Bates, *Language and Context: The Acquisition of Pragmatics*. New York: Academic Press, 1976.

[27] R. Uno and T. Ikegami, "Joint attention / prediction and language: A mechanism to align intentionalities," in *Papers in Cognitive Linguistics* 2, M. Yamanashi, Ed. Tokyo: Hituzi Syobo Publishing, 2003, 231–274. (in Japanese)

[28] J. C. Gomez, E. Sarria, and J. Tamarit, "The comparative study of early communication and theories of mind," in *Understanding Other Minds: Perspectives from Autism*, S. Baron-Cohen, H. Tager-Flusberg, and D. Cohen, Eds. New York: Oxford University Press, 1993.

[29] M. Tomasello, *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge: Harvard University Press, 1993.

[30] K. Onoe, *Grammar and Meaning*, vol.I, Tokyo: Kurosio Shuppan Publishing, 2001 (in Japanese).



Ryoko Uno received her PhD in linguistics from the University of Tokyo in 2006. She is a Senior Assistant Professor in Division of Language and Culture Studies, Tokyo University of Agriculture and Technology.

Her research interests focus on exploring the relationship between bodily movement and grammar from a cognitive linguistic viewpoint.

Davide Marocco received his PhD in Artificial Intelligence at the University of Calabria, Italy, in 2004.
He is currently Lecturer of Cognitive Robotics and Intelligent Systems at the University of Plymouth, UK.

His research interests are focused on Evolutionary robotics models of behavior and evolution of communication and language.

Stefano Nolfi is research director at the Institute of Cognitive Sciences and Technologies of the Italian National Research Council (ISTC-CNR) and head of the Laboratory of Autonomous Robots and Artificial Life.

His research activities focus on Embodied Cognition, Adaptive Behavior, Autonomous Robotics, and Complex Systems. Dr. Nolfi recently edited a book on the "Evolution of Communication and Language in Embodied Agents" published by Springer Verlag.

Takashi Ikegami received his PhD in physics from the University of Tokyo in 1989. Currently, he is a professor at the Department of General System Studies, the University of Tokyo.

His research is centered on complex systems and artificial life, a field which aims to build a theory of life using dynamical systems perspectives.

Prof. Ikegami is a member of the editorial boards of Artificial Life, BioSystems and Interaction Studies.