The Action Game

A computational model for learning repertoires of goals and vocabularies to express them in a population of agents

Bart Jansen¹,² & Jan Cornelis¹

¹ Department of Electronics and Informatics, Vrije Universiteit Brussel, Belgium / ² Department of Future Media and Imaging, Interdisciplinary Institute for Broadband Technology, Belgium

This article introduces a computational model which illustrates how a population of agents can coordinate a vocabulary for goal oriented behavior through repeated local interactions, called “Action Games”. Using principles of self organization and specific assumptions on their behavior, the agents learn the goals and a vocabulary for them. It is shown that the proposed model can be used to investigate the coordination of vocabularies for goal oriented behavior both in a vertical (e.g. between a single teacher and student) and in a horizontal (e.g. in a population of agents which have no a priori repertoire of goals and no vocabularies) transmission scheme. Furthermore, it is shown that in this model shared vocabularies can only be learned from multiple demonstrations consisting of exactly these actions that are strictly required to reach the goal.

Keywords: robot imitation; self organization; vocabularies

1. Introduction

Imagine a teacher or a parent asking a child to try to build a tower. Such an easy request is in fact an impressively difficult task to perform for the child. First of all, it requires the child to understand the meaning of your request. This meaning could be considered as a mapping between “build a tower” and the goal of having several blocks stacked on top of each other (e.g. (Quine 1960)). Even if this meaning is clear to the child, the challenge remains to come up with a concrete plan to achieve the goal, i.e. a series of actions specifically tuned to grasping and stacking the particular objects.
After several interactions, the child will learn to associate the utterance “build a tower” to the goal of having the blocks stacked. This association will evolve during the repeating interactions (Cangelosi et al. 2010). In particular, it will become clear to the child which of the specific object relationships do matter and which are irrelevant (e.g. the blocks must be on top of each other, but it is not relevant which side of the block is at the front side). This learning decreases the probability of failure in future interactions. The process is facilitated by the teacher who is taking extra care to make his point as clearly as possible: the attention of the child to the blocks is drawn by removing other objects from the scene, by gesturing and speech of the teacher etc (Zukow 1990; Zukow-Goldring 1996; Zukow 2006). Once there is shared attention, there will be continuous stimulation, reward and feedback. Feedback could for instance be through language or by helping the child to build the tower.

This paper introduces a computer model of certain aspects of the process sketched above, dubbed the action game. The model illustrates how a vocabulary (i.e. a set of associations between words and goals) for expressing goals can be learned. During the interactions in the form of a game (play), the concepts (the goals) corresponding to these words have to be learned as well. Agents operate in a simple blocks world and can manipulate the blocks. Words correspond to the goal of obtaining a particular spatial configuration of the blocks. Such a configuration can only be achieved by performing a specific sequence of actions that needs to be planned, as the objects in the blocks world are continuously manipulated and changed in position.

After outlining the computational model, three experiments are presented: first, it is shown how a single student can learn from a single teacher, illustrating cases similar to the “build a tower” example. In a second experiment, both the vocabulary as well as the corresponding goals will gradually develop within a larger population and will be shared by all agents in the population. The third part of the paper focuses on two particular aspects of the proposed “Action Games”: the agents learn from multiple demonstrations and the agents need to perform simple and relevant actions for reaching their goals.

The paper is organized as follows: Section 2 discusses related work, the “Action Game” is presented in Section 3. Results of simulation experiments are presented in Section 4 and discussed in Section 5.

2. Related work

Prior to introducing the challenging task of learning goal oriented behavior in robots, the process of learning goals in children is briefly reviewed (Section 2.1). In the discussion section of the paper (Section 5), it is proposed that certain
similarities between some aspects of human goal learning and the proposed computational model might be observed.

2.1 Demonstrating and understanding intentional behavior

The way in which new material is presented to students is important. For instance, the difficulty level of the material compared to the skills of the student (Vygostky 1967), the amount of material, the reinforcement given to the student, etc. all could influence the learning process. Not only teachers are fine tuning those aspects in a classroom setting in order to accommodate the learning process for their students, also parents are subconsciously doing this while interacting with their babies.

One of the best known examples is the way in which mothers adapt their speech while talking to their babies (Snow & Ferguson 1977; Grieser & Kuhl 1988; Fernald & Mazzie 1991; Fernald 1992): their speech becomes high in pitch and contains exaggerated glissando-like rises and falls in pitch. This phenomenon is often called motherese, but is also referred to as Child Directed Speech (CDS) or Infant Directed Speech (IDS) as not only mothers adapt their speech. The function of IDS has been the subject of a debate over the last decades. Several ideas were proposed: according to (Shore 1997) and many others, IDS is important in emotional bonding. Others indicate that the height in pitch makes the sound more appealing (Fernald 1985) and that the modifications are introduced to highlight the relevant parts of the signal (Dominey & Dodane 2004). de Boer & Kuhl (2003) argue that IDS helps the learning of the vowel system of the language. Their conclusion is based on computer simulations in which the learnability of infant directed and adult directed speech is compared. This is in accordance with results from (Liu et al. 2003) showing that the degree to which the mother exaggerates vowels (quantified as vowel space area) correlates to the speech discrimination performance of her child.

Motherese is not limited to the speech domain. It is also observed in caregivers interacting with their deaf children using sign language (Masataka 1992). Mothers also modify their gestures in a similar manner while interacting with their hearing babies (Iverson et al. 1999). Recent findings show that caregivers do not only adapt their language while interacting with babies, but also change the way they act. Similar to motherese, the process of subconsciously adapting actions for demonstrating certain tasks to children is called motionese or infant directed action. When asked to demonstrate a simple task to a baby or to an adult, behavior seems to be significantly different over eight different dimensions of action parameters: range of motion, rate, repetitiveness, proximity to partner, enthusiasm, interactiveness, punctuation and simplification. Differences were found when observers were asked to score the demonstration of mothers to their...
6 to 8 and 11 to 13 month old infants over those eight scales (Brand et al. 2002). Some of the dimensions were also scored automatically by image processing algorithms analyzing video sequences of demonstrations (Rohlًfing et al. 2006). Actions demonstrated by mothers and fathers to 8 to 11 month old infants show more pauses and proceed in a less smooth path compared to adult directed actions (Rohlًfing et al. 2006).

Studies indicate that babies between 1 and 4 months old (Cooper et al. 1997) prefer infant directed speech over regular speech and that infants prefer infant directed sign language over regular sign language (Masataka 1996). Also, infants aged 6 to 8 months and 11 to 13 months prefer infant directed action rather than adult direction action, even if the face of the demonstrator is blurred to obscure emotional and eye-gaze information (Brand & Shallcross 2008). In another study, eight to ten month old infants are found to look longer at movement displays with at least one modified action parameter (range of motion and repetitiveness were studied) compared to static displays (Koterba & Iverson 2009). Infants looking at movements without aspects of infant directed action, do not look longer at it than to a static display. Interestingly, attention is not importantly increased when both range of motion and repetitiveness are modified, compared to only one of them. The same study also shows that infant directed action influences fine-tuned object exploration: low levels of repetitiveness results in infants turning and rotating objects longer than infants exposed to higher levels of repetitiveness. The latter group of infants shows more banging and shaking movements with the objects. It is hypothesized that these movements are actually attempts to imitate the repetitive behavior being demonstrated.

Nagai and Rohlًfing (2008) show that in tasks where the manner in which the task is executed is important, parents highlight the movement of the objects. They also show that in demonstrations of goal oriented behavior, infant directed action can have the effects of (i) highlighting the initial and final states of the action, indicating significant state changes in it, and (ii) underlining the properties of objects used in the action (Nagai & Rohlًfing 2009). Also others have suggested that the main purpose of infant directed action, similar to infant directed speech, is to help the child in recognizing the goal behind an action (Brand et al. 2002). Several studies document these emerging capabilities of children to understand intentional behavior (Meltzoff 1995; Carpenter 1998; Baldwin, D.A. & Baird 1999; Baldwin et al. 2001; Woodward et al. 2001; Olineck & Poulin-Dubois 2005; Buress & Woodward 2007). It was shown that children at the age of 12 months do understand other’s intentions as rational choices of action plans (Schwier et al. 2006), that children expect goals to be achieved in a rational manner (e.g. one-year-olds expect a hand to perform the most direct action available to grab a target object (Woodward & Sommerville 2000)), and that 14 month old...
children try to reach a demonstrated goal by means of the most rational action within their own situational constraints (Gergely et al. 2002). These findings are explained in (Gergely et al. 2002) by the fact that children employ the principle of rational action for discerning intentions behind the demonstrated actions: “The principle of rational action presupposes that (1) actions function to bring about future goal states, and (2) goal states are realized by the most rational action available to the actor within the constraint of the situation.” (Gergely & Csibra 2003, p. 289).

Also robots can benefit from infant directed action for understanding the relevant parts of the actions by using a saliency detection model (Nagai & Rohlfing 2009). For instance, their saliency model can detect pauses between movements, which might be indicative for important states in the task, such as the initial and the end state. Similar to infant directed action, people are found to modify their behavior when tutoring to robots, with respect to hand movement velocity, eye gaze, motion pauses and amplitude (Vollmer et al. 2009).

2.2 Language games

This paper follows the language game approach as advocated by Steels for explaining the emergence of a shared vocabulary in a population of agents. In “The Naming Game” (Steels 1999, 2000; Steels et al. 2002; Steels & Kaplan 2000), agents keep track of a set of weights which express the strengths of the relationships between words and meanings. Given a fixed and shared set of meanings in a population, agents gradually develop a shared vocabulary by continuously adapting these weights: after a successful game, the speaker and the hearer increment the weight of the used “word-meaning” pair and decrement the scores of the competing associations. For the speaker, the competing associations are the pairs with the same meaning (i.e. the topic of the conversation), but with other words. At the hearer’s side, competing associations are those between the same word as the perceived one and other meanings. After unsuccessful interaction, the opposite changes are made (i.e. decrementing the score of the used association and reinforcing the competing associations). This process will be referred to as lateral inhibition. Additionally, mechanisms are introduced which show how the set of meanings can evolve over time or develop starting from an empty set. The focus in the “Naming Game” is on investigating how agents can learn to label each other and can learn to label objects in a scene. These experiments were replicated on fully grounded robots (Baillie et al. 2005). More recently, it was shown how agents can develop a repertoire of simple body motions and learn to label them using the same paradigm (Steels & Spranger 2008a, b). This is similar to (Jansen et al. 2004a) in which repertoires of simple gestures are coordinated based on gestural imitation only.
To some extent the theoretical background underlying the “Naming Game” and the emergence of shared vocabularies has been identified and described in (Lenaerts et al. 2005; Vylder & Tuyls 2006; Baronchelli et al. 2006; Loreto & Steels 2007; Baronchelli et al. 2008).

2.3 Imitation games

The field of robot imitation has advanced in an important manner over the last two decades. Reviews of the field were amongst others given by (Demiris & Hayes 1996; Schaal 1997; Dautenhahn & Nehaniv 2002; Demiris 2007; Nehaniv & Dautenhahn 2007; Argall et al. 2009). In robot imitation, both the learning of meaningless gestures by means of imitation (e.g. (Jansen et al. 2004b; Calinon & Billard 2007; Calinon et al. 2010)) as well as the imitation of goal oriented behaviour was studied (e.g. (Crabbe & Dyer 2000; Billard et al. 2004; Lockerd & Breazeal 2004; Demiris & Khadhouri 2005; Calinon et al. 2005)).

The importance of perspective taking, i.e. the ability to “step into the demonstrator’s shoes”, is common to many of the proposed models for goal level imitation in robots (Trafton et al. 2005; Johnson & Demiris 2005; Breazeal et al. 2006; Demiris & Meltzoff 2008). This perspective taking also plays a key role in human imitation (e.g. the ‘like me’ theory (Meltzoff 2007)). The similarities between both human and robot imitation with respect to perspective taking are described in (Demiris & Meltzoff 2008).

Previously, we have developed a computational model of intention reading in imitation (Jansen & Belpaeme 2006), which fits in the more general multi agent system based framework of imitation games (de Boer 2000; Jansen et al. 2004b) and which was influenced by the language games of (Steels & Kaplan 1999). In this model, one agent comes with a predefined set of goals and wants to transfer those to a second agent. Over multiple demonstrations, the second agent tries to learn the demonstrated goals.

Robot imitation is studied both from the perspective of developmental robotics (Cangelosi et al. 2010), with the objective of increased understanding of human cognition by exploring computational models, as well as from an application point of view; e.g. (service) robots are projected to be introduced in a variety of domains including elderly care. Because robots operating in flexible environments need to be adaptive, they could potentially be instructed to perform new tasks by mere demonstration of the task. Besides adaptivity toward new or changing tasks, natural language interaction is a second requirement for social robots. If natural language processing capabilities are available, robots can learn new tasks from observation, programming by demonstration (Lieberman 1993) and imitation on the one hand and from instructions
in natural language on the other hand. This is explored amongst others by (Nicolescu & Mataric 2001; Mavridis & Roy 2006; Dominey et al. 2009).

3. The Action Game

This section explains the “Action Game” combining aspects both from the “Naming Game” (Section 2.2.) as well as from the “Imitation Game” (Section 2.3). In the “Action Game”, agents develop shared vocabularies of labels for goals, where goals express the agent’s desire to obtain certain spatial relations between blocks, resulting from sequences of actions involving object manipulation. Our model is influenced by the recent work of Steels and Spranger on learning labels for actions (Steels & Spranger 2008a,b) and extends it by learning vocabularies for goals instead of actions, where the goals are not directly observable, but need to be inferred from the observed action sequences. This introduces the additional complexity of goal categorization, which is the core process of inferring goals from sequences of object manipulations.

In the model we propose, agents develop a vocabulary for goal oriented behavior. As in the “Naming Game”, this vocabulary is implemented as a set of associations between words and meanings (goals), where the strength of the association is expressed by a weight. In the “Action Game”, the same learning mechanism based on lateral inhibition as in the “Naming Game” is used for learning the vocabulary (Section 2.2). However, the set of goals in the vocabulary is neither innate nor fixed, but is coordinated through imitative interactions, exactly as in the “Imitation Game” (Section 2.3). Hence, the “Action Game” combines the language learning aspects from the “Naming Game”, with the goal learning aspects of the “Imitation Game” (see also Section 3.5). Although goal oriented behavior can become shared in a population of agents through imitation only (“Imitation Game”), the “Action Game” allows investigating whether linguistic interactions about these goals influence the learning dynamics of the goal learning.

Just as Steels and Spranger do, this paper studies artificial agents (i.e. robots, or computer simulations of robots, called agents), rather than aspects of human task learning. Therefore, while describing the agents, certain skills or properties will be attributed to them which should be considered analogically. For example, agents cannot speak in the sense that they do not have a vocal tract, but there is in the technicalities of the computer simulations a method which can transfer one piece of linguistic information from one agent (referred to as the speaker or the demonstrator) to a second agent (referred to as the hearer). This way, speech production and speech perception are not studied in the model. Similarly, as the
agents do not exist physically, they do not interact physically with their environment and can therefore not perform real actions. Hence, also action performance and perception are simulated and the influence of real life action performance and perception is not studied.

3.1 The agent’s environment and architecture

The environment the agents reside in is the same environment as used in the “Imitation Game” for reading intentional behavior (Jansen & Belpaeme 2006) and consists of an n-by-n discrete two-dimensional blocks board which the agents can observe and manipulate (see Figure 1). The interactions are described in detail in Section 3.2 and always take place between two agents. During these interactions, each agent can observe the blocks world of the other agent. Each agent is equipped with four action primitives, i.e. it can move the blocks on the board one cell in each of the four directions. Left(A) will be used to refer to the action of moving a block identified as A one cell to the left in the blocks world. The agents can perceive the blocks on the board and they can reason in terms of spatial relations over the blocks. In the extremely simple simulated environment studied in this paper, agents only recognize two spatial relations: they know whether a block is above another one or whether a block is at the left of another block. In this simple blocks world, the goals learned and represented by the agents are used to obtain a blocks world which has certain desired properties, expressed as a combination of these spatial relations. For instance, the goal Above(A, B) ∧ LeftOf(A, C) expresses the desire of the agent to obtain a blocks world in which block A is above block B and block A is at the left of block C. As pointed out by (Searle 1984), the difficulty in discerning goals in a sequence of actions is that goals can typically be pursued by performing different action sequences and that an action sequence can pursue different goals. Even in this very simple blocks world, a one-to-one mapping between action sequences and

![Figure 1. A two-dimensional blocks world in which the goal Above(A, B) ∧ LeftOf(A, C) holds](image-url)
goals does not exist and hence, it is impossible for the hearer to learn the goal of the speaker’s act from a single demonstration and without additional information (Jansen & Belpaeme 2006).

The agents are equipped with a structure, called repertoire, which stores the learned goals, together with a score expressing how often the usage of the goal leads to a successful game and a usage count. Besides the repertoire of goals, agents also have a vocabulary of words they can use in communication. The vocabulary is a structure containing triples consisting of a word, a goal and an association score, bound between 0 and 1.

The motor planning aspects of action performance in the real world, nor the visual processing of action perception are investigated in this article. Therefore, it is assumed that (1) all agents can perform the same four action primitives, that (2) all agents can observe these four actions being executed by the agents they interact with and that (3) all agents can categorize the performed action sequences in terms of combinations of the same spatial relations. Under these restrictions, it is investigated how shared repertoires of goals and shared vocabularies can be learned through imitation.

### 3.2 The interaction pattern

The “Action Game” is an interaction between two agents; one will take the role of speaker, the other agent will take the role of hearer. Several interaction patterns are possible. One could for instance study a vertical transmission scheme (Cavalli-Sforza & Feldman 1981). This is the case where one agent is an experienced teacher (i.e. with a repertoire of goals and a vocabulary for it) and another agent, which is the student, that has to acquire the repertoire or vocabulary (see Section 4.1). Alternatively, one could study populations in which all agents start without repertoires of goals and vocabularies and have to develop these during interactions. Experimental results of our computational model employed in such a horizontal transmission scheme (Cavalli-Sforza & Feldman 1981) are reported in Section 4.2.

Figure 2 shows the general outline of the game. At the start of each game, both agents rearrange the blocks in their blocks world to random positions. The game starts when the speaker selects a random goal from its repertoire. In its vocabulary, it looks up the word for this goal and utters the word. The hearer hears the word and finds the goal for this word. The hearer builds a plan for reaching this goal (Section 3.3) and performs these actions. This sequence of actions is observed by the speaker who categorizes this sequence of actions as the goal best explaining the observed action sequence (Section 3.4). If this goal is equal to the initial goal selected by the speaker, the game succeeds. In all other
cases, it fails. The speaker signals to the hearer the outcome of the game. If the game fails, the speaker demonstrates to the hearer how to obtain the intended goal, without expecting a reply from the hearer. Based on the outcome of these games, both agents adapt their repertoires of goals and their vocabulary in order to gradually reach shared vocabularies (Section 3.5). If the hearer can categorize this observed sequence of actions as a goal, he can associate the word used by the speaker to this goal. Worked out examples of “Action Games” are provided in Section 3.6.

As can be observed in Figure 2, goals themselves are of course never exposed directly to the agents. Either, there is linguistic interaction (lower part of the figure), where the agents need to learn the mapping between the goals and words. Either, the agents interact by performing and observing sequences of actions (upper part of the figure), where the challenge is to generate action sequences for goals and to interpret action sequences as intentional behavior, i.e. as goals. The repertoire
of goals itself is not given nor fixed, it has to be constructed, together with the vocabulary and the bi-directional mappings.

3.3 Planning

In the majority of cases, the agent cannot reach the targeted goal by performing a single action. Hence, a plan must be built which specifies the action sequence that must be executed to reach the desired goal state, starting from the current configuration of the blocks. In our framework, it is assumed and ensured that a plan is always optimal; this means it consists of the minimal number of actions required for reaching the goal starting from the current configuration. Typically, multiple optimal plans exist, in that case one of the optimal plans is chosen at random. In the computer simulations, the optimal plan is found by using the A* algorithm (Hart et al. 1968) with an admissible heuristic which is estimating the minimal number of horizontal moves required to reach each $LeftOf$ predicate in the goal, summed to the minimal number of moves required to reach each $Above$ predicate occurring in the current goal.

Hence, in this experiment, the agents have a fixed and innate planning mechanism, which maps goals onto sequences of actions. This mapping is not a 1-to-1 mapping and does not have to be the same for all agents (Jansen & Belpaeme 2006).

3.4 Goal categorization

When the game fails, the speaker demonstrates how to reach the intended goal (visualized as step 3 in Figure 2). When the hearer observes this sequence of actions being performed by the speaker, it has to interpret these actions as pursuing a particular goal. The hearer does this by maintaining a memory (the repertoire) of goals already experienced before. Associated to every goal, a score is kept. Upon every demonstration, the hearer inspects his repertoire for a goal which matches the observed sequence of actions. From all goals which are relevant and optimal, the goal with the highest score will be the goal he uses to describe the speaker’s action sequence with.

The hearer’s inspection of its repertoire for a goal matching the observed sequence of actions is guided by two criteria: relevance and optimality.

1. A goal is relevant for the observed sequence of actions if the goal is reached after but not before the last action executed.
2. A goal is optimal for the observed sequence of actions if the optimal plan for the goal has the same length as the observed sequence of actions.
It is important to note that it is not straightforward for the hearer to test for optimality: for every goal in its repertoire, it has to take the perspective of the speaker and imagine (simulate) the plan he would have built for this goal starting from the speaker’s initial blocks configuration. This “perspective taking” is computationally expensive, but does provide the hearer with a clear insight in the reasons why the speaker performed the observed action sequence.

3.5 Learning

For each game, the speaker announces to the hearer whether it is successful or not. Given this feedback, together with the observed action sequence performed by the speaker in case of a failing game, the hearer has to learn the goals and words the speaker tries to transfer. This learning is operationalized by maintaining the repertoire of goals and a vocabulary. The following learning operators are defined for maintaining the repertoire of goals:

1. **The construction of new goals**: When the goal categorization process fails, it means that the hearer cannot identify a goal in his repertoire compatible with the speaker’s action sequence. In that case, the hearer creates a new random goal which is optimal and relevant for the observed sequence of actions and adds it to its repertoire, so that whenever he observes the same goal being demonstrated again, the goal categorization process might now succeed. The same mechanism as proposed in the “Imitation Game” (Jansen & Belpaeme 2006) is used here and involves several steps. First, all possible relevant goals are enumerated (i.e. all those goals which are reached in the last but not in the second last state). For each of these goals, it is evaluated whether they are optimal, i.e. the hearer checks for each of these goals whether he would need the same amount of actions to reach these goals (starting from the initial board’s configuration of the speaker), as the speaker needed to reach the currently demonstrated goal.). From all relevant and optimal goals explaining the observed sequence of actions, the hearer selects the shortest goal (i.e. the goal represented by the least amount of predicates) and adds this goal to its repertoire.

2. **Adaptation of scores of goals**: When the interaction is a success, the score of the goal used by the hearer is increased with an amount $\delta_{goal}$. The scores of the other goals which were found to be relevant and optimal during the categorization process are decreased by an amount $\delta_{goal}$. When the game fails, the opposite is done. In any case, the score of a goal is limited between zero and one. The score of a goal which is added to the repertoire is initialized at 0.5.
3. **Removal of goals**: Whenever the score of a goal drops below a threshold $\varepsilon_{goal}$, it is removed from the repertoire of the hearer, as it has shown not be useful for categorizing the observed sequences of actions.

The following learning operators are defined for maintaining the vocabulary:

1. **The construction of new words**: When the vocabulary of the speaker is empty, a new word is added to it. New words are unique symbols, typically random combinations of syllables. When the hearer does not know the word he’s hearing, he adds it to his vocabulary and associates it to the recovered goal with an initial association strength of 0.5.

2. **Adaptation of the vocabulary scores**: If the game succeeds, a lateral inhibition mechanism increases the score of the used word-goal pair with an amount $\delta_{voc}$ but decreases the score of all competing associations in both agents with the same amount. If the game fails because of the goal demonstrated by the hearer is not the goal intended by the speaker, both agents decrease the score of the used word-goal pair, but increase the scores of the competing pairs.

3. **Removal of words**: When the association strength between a word and a goal drops below a certain threshold $\varepsilon_{voc}$, the association is removed from the vocabulary.

3.6 Example games

Below, two example games are provided to illustrate the specifics of the goal categorization process.

**GAME 0** GAME 0 is the very first game (i.e. the first interaction) between a single speaker and a single hearer. Suppose the speaker has a simple blocks world (shown in Figure 5 at the very left), and a vocabulary with a repertoire of goals as shown in Figure 3. The hearer has not acquired any goal or item in its vocabulary. In this first example, the blocks world of the hearer is not relevant yet. Assume that the speaker initially selects the goal $LeftOf(A, B)$ from its repertoire. Looking in its repertoire of goals, the speaker finds the word “waba” associated to it. So, the speaker utters the word “waba”. As the vocabulary of the hearer is empty, the hearer cannot identify the goal and the game fails. Upon failure, the speaker demonstrates how to reach the goal. Therefore, the speaker needs to build a plan. For reaching the goal $LeftOf(A, B)$ starting from the board as shown in Figure 5, four different relevant and optimal plans exist: $Right(B) \rightarrow Right(B)$, $Left(A) \rightarrow Left(A)$, $Right(B) \rightarrow Left(A)$ and $Left(A) \rightarrow Right(B)$ ($X \rightarrow Y$ denotes “$Y$ is performed after $X$”). Suppose the speaker selects the plan $Left(A) \rightarrow Left(A)$ and hence moves block $A$ two cells to the left, see Figure 5.
Meanwhile, the hearer observes this action sequence as performed by the speaker and needs to interpret these actions as goal oriented behavior. The repertoire of goals of the hearer is still empty and hence no goal in its repertoire can explain the observed sequence of actions. By consequence, goal categorization fails. The hearer therefore constructs a new random goal which is optimal and relevant and associates it to the word “waba”. Note that there are in general several goals which are optimal and relevant for the observed sequence of actions: Besides the true goal of the speaker $\text{LeftOf}(A, B)$, also $\text{LeftOf}(A, C)$ is optimal and relevant in this case. From the current demonstration, the hearer has no means to prefer any of the two goals. So, it is possible that the hearer associates the wrong goal with the current word. We suppose that the hearer associates $\text{LeftOf}(A, B)$ to “waba”; its vocabulary after game 0 is illustrated in Figure 4.

<table>
<thead>
<tr>
<th>Goal</th>
<th>Waba</th>
<th>Paka</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{LeftOf}(A, B)$</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>$\text{Above}(A, B) \land \text{LeftOf}(B, C)$</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 3.** The vocabulary of the speaker at the beginning of GAME 0

<table>
<thead>
<tr>
<th>Goal</th>
<th>Waba</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{LeftOf}(A, B)$</td>
<td>0.5</td>
</tr>
</tbody>
</table>

**Figure 4.** The vocabulary of the hearer at the end of GAME 0

GAME 1 Before starting the game, the speaker and the hearer reshuffle the blocks in the environment. Suppose the hearer starts the game with the blocks world shown in Figure 6. Suppose that in GAME 1, the speaker again selects the goal $\text{LeftOf}(A, B)$ and hence says “waba”. Upon hearing this word, the hearer now can retrieve the goal $\text{LeftOf}(A, B)$ from its repertoire. In order to reach this goal, the hearer builds a plan (e.g. $\text{Up}(A) \rightarrow \text{Left}(A) \rightarrow \text{Left}(A)$) and executes it. The speaker observes this sequence of actions and correctly identifies this action sequence as $\text{LeftOf}(A, B)$, as this is the only goal in its repertoire which is relevant and optimal for the observed action sequence. Hence, this game succeeds. Note that the speaker also has the goal $\text{Above}(A, B) \land \text{LeftOf}(B, C)$ in his repertoire which is also reached after the sequence of actions performed by the hearer. This goal is not relevant and optimal, given the fact that it is already reached after the first action ($\text{Up}(A)$) and hence the entire action sequence is not required. Note that there are cases where the game fails because the speaker does not correctly categorize the actions performed by the hearer, even if the hearer pursued the correct goal.
4. Experiments

4.1 Single speaker and single hearer (vertical transmission scheme)

In this experiment a single speaker interacts with a single hearer. The speaker starts with a vocabulary of \( k \) different random words, associated one to one to a repertoire of \( k \) different random goals. The hearer starts without a repertoire of goals and without a vocabulary. This experiment investigates whether the vocabulary and the repertoire of goals can be transmitted to the single hearer by repeated “Action Games”.

In this experiment, the board size of both agents is 5-by-5 cells, containing three blocks (\( A, B \) and \( C \)). The agents both can perform the same four actions (\( \text{Left}(X) \), \( \text{Right}(X) \), \( \text{Up}(X) \) and \( \text{Down}(X) \) where \( X \in \{A,B,C\} \)) and perceive the blocks world in terms of the spatial relations \( \text{LeftOf}(X, Y) \) and \( \text{Above}(X, Y) \) where both \( X \) and \( Y \) can be \( A, B \) or \( C \), but cannot be equal to each other. Goals are either single relations (e.g. \( \text{Above}(A, B) \)) or a conjunction of two relations (e.g. \( \text{Above}(A, B) \land \text{Above}(B, C) \)). \( \delta_{\text{goal}} \) was set to 0.025, \( \varepsilon_{\text{goal}} \) was set to 0.1, \( \delta_{\text{voc}} \) was set to 0.1 and \( \varepsilon_{\text{voc}} \) was set to 0.2. These parameter settings are in accordance with the parameters used in the “Imitation Game” simulations (Jansen & Belpaeme 2006).
and in the “Naming Game” simulations (e.g. (Steels 1999)). The experimental results are obtained by averaging over ten runs of 10000 “Action Games” played between the two agents. Confidence intervals illustrate the variations among ten runs. The graphs show a running average over 1000 games in order to obtain smooth graphs. The number of goals and words in the vocabulary of the speaker is varied $k \in \{10, 20, 50\}$.

The upper graph in Figure 7 shows the fraction of successful games. After 2000 games, the success rate is $0.96(0.29)$, $0.87(0.12)$, $0.73(0.43)$ for initial repertoires of 10, 20 and respectively 50 goals. Numbers between brackets are standard deviations. After 10000 games, the success rate raises to $0.99(0.12)$, $0.96(0.08)$ and $0.79(0.38)$. So, the agents reach nearly 100% success rate for 10 and 20 goals, but for 50 goals the success rate remains suboptimal. When more and more “Action Games” have taken place, the standard deviation of the success rate decreases. Throughout the game sequence, the standard deviation of the success rate is higher for larger number of goals in the speaker’s repertoire. Similar trends are observed with respect to the number of words in the vocabulary of the hearer (Figure 7, middle graph) and with respect to the goal repertoire size (Figure 7, lower graph). From the three graphs it is concluded that the more goals and words in the initial vocabulary of the speaker, the harder it is to transmit these to the hearer.

The same trend as the one described above is observed in vertical transmission experiments in the “Imitation Game” with various repertoire sizes (e.g. (Jansen & Belpaeme 2006)) and can be explained by the fact that for large repertoires, the probability that certain goals in the repertoire are subsets of other goals is high (e.g. $\text{Above}(A, B)$ against $\text{Above}(A, B) \land \text{Above}(B, C)$). As there is a preference in the goal categorization process for more general goals over more specific goals, several specific goals sometimes tend to be replaced by a single more general goal. This is also observed in the presented simulation experiment; after 200000 games the average goal repertoire sizes of the hearer are $11.99(1.88)$, $19.82(2.57)$ and $33.23(3.66)$ for respectively a speaker starting with 10, 20 or 50 words and goals in his repertoire.

Further comparison between the results obtained from the “Action Game” and from the “Imitation Game” (Jansen & Belpaeme 2006) with the same parameter values in a vertical transmission setting for various initial repertoire sizes of the demonstrator reveals that there is very little difference between the evolution of the success rate and the goal repertoire size: for instance, after 2000 games the success rates in an “Imitation Game” with the same parameters are $0.95 \ (0.04)$, $0.86 \ (0.06)$ and $0.76 \ (0.06)$ for 10, 20 and respectively 50 goals in the repertoire of the demonstrator. In the “Imitation Game”, repertoires of goals are learned based on imitative interactions only; whereas in the “Action Game” linguistic information...
and observed behavior are combined to learn the repertoires of goals. In the vertical transmission case, the introduction of the additional mapping between goals and words which is to be learnt compared to the “Imitation Game”, does not alter the successful acquisition of the repertoire of goals.
4.2 A population perspective (horizontal transmission scheme)

In this second experiment, a population of $k$ agents starts without a vocabulary and without repertoires of goals. The agents have to coordinate a shared vocabulary and shared goals through local interactions only. In this horizontal transmission scheme, all agents can take the role of hearer and speaker. At the start of each game, two agents are randomly selected from the population and are randomly assigned the roles of speaker and hearer. In this manner, over the course of many games, all agents will interact with all other agents several times, both in the roles of speaker and hearer. The same experimental parameters and scores are used as in the vertical transmission experiment described in the previous section. Compared to the vertical transmission scheme, there is a single difference in the game: in the horizontal transmission scheme, both the speaker and the hearer can add a new random goal to their repertoire of goals with a small probability $\gamma$, which is set to 0.1. This drives the agents towards exploring the entire space of goals. Figure 8 shows results for an experiment in which the number of agents $k$ in the population is varied ($k \in \{10, 20, 50\}$). The provided results are averages over all agents in the population. In the upper graphs the success rate is plotted. It clearly shows that it takes longer before the success rate stabilizes (and at a lower rate) for larger population sizes. For instance, after 50000 games, the success rates are 0.84(0.33), 0.60(0.48) and 0.30(0.45) respectively for populations containing 10, 20 and 50 agents. After 200000 games, the success rates reached 0.95(0.13), 0.90(0.23) and 0.61(0.48) respectively.

The middle graph in Figure 8 shows the number of words created by the agents. An initial “overshooting” phase is observed, where there are 2 to 5 times more words than goals. This is caused by synonymy, because several pairs of small groups of agents within the population can associate different words to the same goal, as agents invent random words for the goals they cannot express. However, after a sufficient amount of games, the lateral inhibition mechanism operating over the vocabulary will result in a winner-take-all effect: as soon as a single word gets successfully used a little bit more often for expressing a given goal, its score will be increased more and more. Hence, the agents will more often use that word for expressing the given goal. In the end, the synonyms for this word in the vocabularies will not remain successful enough to be kept in the vocabularies of the different agents.

The lower graph in Figure 8 shows the size of the repertoire of goals learned by the agents. The standard deviations of the number of goals over the 10 runs is so small that the 95% confidence intervals are not visible on the graph. In each of the 3 experiments studied ($k = 10, 20, 50$), the goal repertoire size raises to 72 goals. In the given experimental settings, there are 72 different goals which can be reached
with one or two relations over three blocks. Hence, agents succeed to cover the entire goal space and to label each goal successfully.

A comparison between a population of 10 agents playing “Imitation Games” and a population playing “Action Games” was performed and is shown in Figure 9.
After 200,000 games success rates of 0.77(0.41) and 0.90(0.23) for the “Imitation Game” and the “Action Game” are observed, while the repertoires of goals contain respectively 54.35(1.07) and 72(0) goals on average. In the “Action Game”, agents better succeed in developing categories over the entire goal space, compared to the “Imitation Game” (72 against 54), also resulting in higher success rates (0.90 against 0.77). This is explained by the fact that the lexical interactions in the “Action Game” inhibit the preference for shorter over longer goals in the categorization process. The latter results in several specific goals to be categorized as a single more general goal in the “Imitation Game”. In the “Action Game”, the different words used by the speaker help the hearer in discriminating between the more general and more specific goals which could not always easily be separated in the “Imitation Game”.

4.3 A modified goal categorization method

The action sequences which the speaker performs are assumed to be as simple as possible for demonstrating the selected goal. In other words, it is assumed
that the agents only execute these action sequences directly relevant for obtaining the goal. The model implements this by assuming that the plan constructed by the speaker must be optimal in the number of actions. The hearer explicitly uses this criterion for filtering candidate goals from the enormous space of possible goals for a given observed sequence of actions. In an additional experiment the influence of the optimality of goals is investigated in a population of agents which have a modified goal categorization mechanism that categorizes an observed sequence of actions as any relevant goal, rather than as a relevant AND optimal goal.

Figure 10 shows results of simulation experiments with a population of 10 agents, with modified goal categorization, playing “Action Games”. After 200000 games, the success rate is 0.47(0.50), which is much lower than 0.90(0.23) as observed when 10 agents interact in “Action Games” using the same parameter values, but with the non modified goal categorization process, as shown in Figure 8. Even after 1 million games, the vocabulary size remains much bigger than the number of goals and the success rate remains below 80%.

![Figure 10](image)

**Figure 10.** The evolution of the success rate, the vocabulary size and the repertoire size over time for “Action Games” in which the hearer does not assume that the speaker is performing the most rational action sequence. Although repertoires of goals and vocabularies are learned, the success rate is very low and increases very slow.

5. Discussion

The “Action Game” is not a model of one-shot learning, neither are the “Naming Game” and the “Imitation Game” on which the “Action Game” is built. The experimental results reported in Section 4 illustrate that the goal repertoire size, the vocabulary size and the success rates stabilize after several thousands
of games in the vertical transmission case. In the horizontal transmission case, stabilization only occurs after several tens of thousands of games. Multiple games about the same goal, each starting from a different configuration of the blocks in the 2D blocks world are crucial in order to sort out the ambiguous relation between externally observable action sequences and internal goals. By recognizing the regularity in the speaker’s behavior over multiple actions, the hearer can correctly deduce the intentions behind the speaker’s behavior. Also, preferences over goals can be learned over multiple demonstrations by adapting the scores associated to each goal. In the horizontal transmission case, due to random fluctuations, some smaller subgroups of the population tend to develop their own words for some goals, which results in synonymy. In order to converge to a clear winner-take-all situation with a single word associated to each goal, even more interactions (about a wide variety of goals) are required between all different pairs of agents.

The basic idea of action gaming is using the invariant features (of observed behaviour on top of linguistic information) in multiple demonstrations under different conditions of the same task for understanding the goal behind it. This idea is also employed outside of the context of “Action Games” in many other studies investigating robot interactions and learning, e.g. (Sato et al. 2003; Ekvall & Kragic 2006; Pardowitz et al. 2007). The number of games required to convey a repertoire of goals and the vocabulary from a speaker to a hearer in our model, exceeds by far the number of explicit interactions in real life between parent and child. This is explained by the fact that human interaction is far more complex than the minimalistic agent interactions presented here. Various additional mechanisms not studied in the “Action Game” constrain the search space for the child and include amongst others shared attention, object affordances (Montesano et al. 2008), more advanced perspective taking, additional social parental feedback and so on. However, it was observed that mothers repeat demonstrations more often when demonstrating a task to their children compared to demonstrations to adults (Brand et al. 2002) (see Section 2.1 for an overview). We strongly believe that this repetitiveness allows the child to identify the invariant aspects over multiple demonstrations of the same goal and that this is required to infer the goal behind sequences of actions, i.e. to gradually refine the goal categorization over multiple interactions. In the “Action Game” presented above, goal categorization is refined over multiple games exactly by observing the invariant features. This is also observed in visual attention and saliency models (Nagai & Rohlfsing 2008) which amongst others are triggered by repetition and help robots in detecting meaningful segments in sequences of actions and derive goals from this.

Besides the crucial role of repeated demonstrations for reaching a shared vocabulary for goal oriented behavior in a population of agents playing “Action
Games”, a second key property of the “Action Game” is that the hearer must assume that the sequences of actions which are performed by the speaker when demonstrating a goal are both relevant and optimal. Without this assumption the population of agents fails to develop a shared vocabulary for the goals, as was shown in Section 4.3. Even in the very simple discrete and small blocks world with very few actions and simple spatial relations, the amount of possible goals is enormous. From this enormous set of possibilities, the hearer has to identify the correct goal. The requirement that goals explaining a sequence of actions are optimal, i.e. are performed in the least amount of required steps, reduces the size of this hypothesis space in an important manner.

The principle of optimality employed by the hearer in the goal categorization process in the “Action Game”, shows some similarity to the principle of rational action (Gergely & Csibra 2003) which states that children assume that the goal state is realized by the most rational actions available to the actor. This assumption made by the child is compatible with some of the behavior changes made by parents in infant directed action. In particular, parents subconsciously modify the demonstrated act (Brand et al. 2002).

In the computational model we have proposed, agents can only learn a shared repertoire of goals and a vocabulary for them if they explicitly can assume that goals are relevant and optimal and if they can rely on multiple games for identifying the goals behind the observed actions. These assumptions do not help in making the demonstration more appealing to the learner, in emotional bonding or in understanding object affordances as these aspects were not incorporated into the “Action Game”. They explicitly reduce the search space of possible goals explaining an observed act. We have shown that without these assumptions built in the behavior of the agents, they fail to develop a shared vocabulary for labeling goal oriented behavior.

Cangelosi (2010) reviews three different models with increasing complexity of the development and acquisition of linguistic capabilities based on symbol grounding. Most existing work in the field is described as a model of direct grounding, i.e. grounded symbols are acquired by the agents through embodied interactions with the environment and are grounded in perceptual or sensorimotor interactions with this environment. The model we proposed in this article, is a model in which each symbol (a word) is directly linked to a grounded concept (a goal), be it in a less straightforward manner compared to objects or actions. Cangelosi (2010) argues that only a minimal kernel of a language is in fact grounded directly and that meanings are grounded in a mental lexicon that is in turn grounded in prior sensorimotor learning. However, in particular for goal oriented behavior, further research is required to investigate the complex interplay between language and action, also after the initial acquisition of grounded symbols.
6. Conclusion

This paper illustrates how a population of agents can establish a shared repertoire of goals and a vocabulary for them. Although goals cannot be directly observed in the proposed “Action Games”, the vocabulary and the repertoires of goals are shared after multiple games. The “Action Game” combines aspects from the “Naming Game” for learning vocabularies in a population of agents with aspects from the “Imitation Game”, a computational model of intention reading in imitation. In the “Imitation Game”, repertoires of goals are learned based on imitative interactions only.

Experimental results show that the linguistic interactions about the goals in the “Action Game” improve the sharing of the repertoires of goals among the agents, compared to imitative interactions only. This allows agents to engage in a higher amount of successful games and to learn bigger repertoires of goals.

References


Authors’ addresses

Jan Cornelis
Department of Electronics and Informatics
Vrije Universiteit Brussel
Pleinlaan 2
1050 Brussel
Belgium

jpcornel@etro.vub.ac.be

All rights reserved
Bart Jansen  
Department of Electronics and Informatics  
Vrije Universiteit Brussel  
Pleinlaan 2  
1050 Brussel  
Belgium  

Department of Future Media and Imaging (FMI)  
Interdisciplinary Institute for Broadband Technology (IBBT)  
Gaston Crommenlaan 8 (box 102) B-9050 Ghent  
Belgium  

bjansen@etro.vub.ac.be

Authors’ biography

Bart Jansen obtained his Ph.D. in computer science in 2005 with a dissertation on robot imitation. Since then, he’s active as a post doc researcher and professor at the Engineering Science Faculty of the Vrije Universiteit Brussel, Belgium. Although his research currently focuses on aspects of artificial intelligence in e-health and telemedicine, he remains interested in the domains of language evolution and robot imitation.

Jan Cornelis, 17/10/1950, MD-1973, Ph.D.-1980. Professor in digital image processing and electronics, coordinator of research group IRIS (computer vision, image processing), academic coordinator Knowledge Innovation and Technology Transfer at VUB. Current research interest: Image and video compression, Medical imaging. R&D-management experience: co-founder of BI3 NV incubation fund, and ICAB NV – Incubation Centre of VUB. In 2001–2008 he was Vicerector for research at VUB. Afterwards, he became deputy head of cabinet in the Ministry of Science Policy and Innovation. He was representing the Flemish Government in the Board of Directors of IMEC.