Emergent Mirror Systems for Body Language

Luc Steels and Michael Spranger

This paper is the author's draft and has now been officially published as:

Steels, Luc and Spranger, Michael (2012). Emergent mirror systems for body language. In Luc Steels (Ed.), *Experiments in Cultural Language Evolution*, 87 – 109. Amsterdam: John Benjamins.

Abstract

This chapter investigates how a vocabulary for talking about body actions can emerge in a population of grounded autonomous agents instantiated as humanoid robots. The agents play a Posture Game in which the speaker asks the hearer to take on a certain posture. The speaker either signals success if the hearer indeed performs an action to achieve the posture or he shows the posture himself so that the hearer can acquire the name. The challenge of emergent body language raises not only fundamental issues in how a perceptually grounded lexicon can arise in a population of autonomous agents but also more general questions of human cognition, in particular how agents can develop a body model and a mirror system so that they can recognize actions of others as being the same as their own.

1. Introduction

Commands requesting the hearer to take on certain bodily postures, such as "siting down", "standing up", "turning around", "your left arm raised", are among the most basic words in a language and they are acquired at a very early age. Nevertheless, their origins and acquisition raises a number of deep open problems in cognitive science:

1. *Body Image*: Language users need a model of their own body, which includes a distinction between different bodily components (arms, legs, hands), a motor body model to track bodily movements, a visual body model to relate perception with action, and a simulation model to predict what the effect of an action is going to be. How do such body models operate and develop?

2. *Mirror systems*: Language users must be able to relate the perception of body postures by *another* person to their own body postures, otherwise the speaker would not be able to see whether the hearer correctly executed the demanded action and the hearer could not learn the body posture associated with an unknown name. How does such a mirror system work and how does it arise?

3. *Symbol Grounding*: Language users must reach an agreement on which names to use for actions, without telepathic direct meaning transfer and without a central coordinator. These names must be grounded in motor programs to execute the action and visual recognition routines to recognize the action.

The subject of body image has received wide attention, particularly in the neurological literature, because of puzzling phenomena such as phantom limbs, mirror box experiments, unusual pain, out of body experiences, etc. (Ramachandran & Hirstein, 1998; Rosenfield, 1988; Blanke & Castillo, 2007). The discovered disorders, experiments and neurobiological observations make it quite clear that body image is not a simple, fixed innately given internal representation but a dense network of representations (Ghahramani & Wolpert, 1997) that forms in development and continues to change and is adapted throughout life in order to handle growth, aging or change due to accidents (Blanke & Castillo, 2007). Even in robots, the behavior of motors, but also the morphology of a robot might change over time and the body model must be constantly adjusted Hoffmann et al., 2010.

The field of neuroscience and neurobiology has also focused on body image lately and particularly on the mirror system problem because of the discovery of mirror neurons. Mirror neurons are single neurons which are active both when an individual performs a particular action, but also when the individual perceives the action performed by others. The discovery of such neurons in primates (Rizzolatti et al., 2001) and corresponding systems in humans (Rizzolatti, 2005) have led to widespread hypotheses about their role in action-understanding, imitation and intention-understanding (Gallese et al., 2004), as well as conceptual knowledge (Gallese & Lakoff, 2005) and even language (Rizzolatti & Arbib, 1998; Pulvermüller, 2005).

Moreover, profound differences between human languages exist in terms of which actions are named and how the body is conceptualized for language (Talmy, 2000). For example, several languages, including Serbo-Croatian, do not lexicalize the distinction between arm and hand, foot and leg, finger and thumb, or finger and toe. These cultural differences suggest that there must be a strong learning component and cultural influence in how the body is conceptualized for language. But this raises a chicken-and-egg problem: body language requires a grounded mirror system but it influences itself what action concepts such a mirror system should support.

The relation between visual representations and recognition of bodily action on the one hand and the bodily action itself has already been intensely studied in robotics research, particularly in research on imitation (Billard, 2002; Demiris & Johnson, 2003). It is moreover a key topic in 'embodied artificial intelligence' (Pfeifer & Bongard, 2006) which emphasises the grounding of cognition in bodily activities. It is currently well established that acquiring relations between visual and motor body image is a very non-trivial problem. The present chapter argues that a *whole systems approach* is essential for solving the problem. All components of the semiotic cycle needed for a complete verbal interaction (perception, conceptualization and sentence production for the speaker and parsing, interpretation and motor action for the hearer) need to cooperate and co-evolve and should be tightly integrated so that one component can help to bootstrap another one.

We will use a language game called the *Posture Game* in which one robot asks another one to achieve a particular body posture, such as "right arm raised". The game is a success if the hearer indeed carries out an action that achieves the requested posture (Figure 1). The game has been operationalized on non-commercial Sony humanoid robots (Fujita et al., 2003) and is described in more detail in (Steels & Spranger, 2008a,b, 2009; Spranger & Loetzsch, 2009). Although there are only two robot bodies available, a population is simulated by storing the agents' internal states as software states on a server and loading these states into each robot at the beginning of a game. At the end of the game, the internal state is uploaded back to the server. This kind of facility makes it not only possible to have larger populations with only a limited number of robot bodies, but also to do repeatable experiments by recording and re-using the sensory-motor data of embodied experiments.

A single Posture Game works as follows. Two agents are randomly chosen from the population and downloaded in two robot bodies. The robotic agents then go through the following script:

1. The speaker chooses a posture from his inventory of postures.

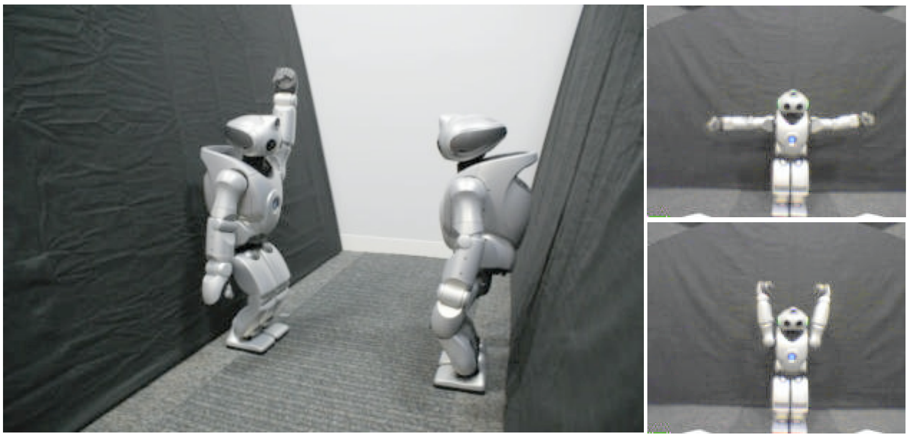

Figure 1. *The experimental setup, where two humanoid robots face each other and play a Posture Game, in which the speaker asks the hearer to take on a certain posture, like "left arm raised", or "both arms stretched above the head". The right image shows two example postures as seen through the camera of another robot. Agents start without any prior body model, knowledge of possible actions, or names for them and have to autonomously bootstrap a body language system from scratch.*

2. The speaker retrieves the name for this posture in his vocabulary and transmits that to the hearer.
3. The hearer retrieves the posture by looking up the name in his own vocabulary and evokes the motor behavior that could achieve this posture.
4. The speaker observes the posture adopted by the hearer and checks whether it fits with the prototypical visual body-image of the posture he had originally chosen.
5. If this is NOT the case, the speaker signals failure. The speaker activates his own motor behavior for achieving this posture in order to repair the communication, so that there is an opportunity for the hearer to learn the speaker's name for this posture.
6. Otherwise the speaker signals success.

We will proceed in steps, at first scaffolding some aspects of the problem, until agents can play autonomously the Posture Game. In the first experiment (the mirror experiment) the robot acquires the relation between body postures and actions through kinesthetic teaching and the relation between body postures and visual appearances by observing itself before a mirror. The second experiment removes the use of a mirror, showing that the emergent language can itself help the robots to acquire and coordinate the relation between a bodily action and its visual appearance. The final experiment adds the capacity of self-simulation, which helps to constrain the space of possible meanings of an unknown command.

2. The mirror experiment

Body image refers to a collection of representations that embodied agents must maintain in order to move about in the world, plan and execute action, perceive and interpret the behaviors of others, build and use an episodic memory, and understand or produce language about action, for example commands.

2.1. Acquiring a Mirror System

The Mirror Experiment has been designed as a first step to investigate how robots could develop effective body models prior to language. Kinesthetic teaching is used to endow the robot with a motor control program to achieve a particular posture. Kinesthetic teaching means that the experimenter brings the robot manually into a particular posture. The robot records the critical angles needed to reconstruct the motor control program, the expected proprioception and other additional feedback loops and information needed to robustly achieve the posture.

Next each robot stands before a mirror in order to acquire the relation between its own motor body-image and (a mirror image) of visual body-image (see Figure 2). The robot performs motor babbling, choosing randomly motor control programs from the repertoire acquired with kinesthetic teaching. In other experiments, robots simply look at their own body (Figure 3 left) or use a smaller mirror to inspect parts of the body which they cannot see easily otherwise (Figure 3 right). Once the robotic agents have acquired a repertoire of possible body postures, which motor programs are needed to achieve them, and what they look like, the same repairs and alignment strategies as used in the Naming Game can be used to self-organize a vocabulary.

We propose that the internal body image representations of a robot are organized in terms of a *semiotic network* that links sensory experiences, posture proto-

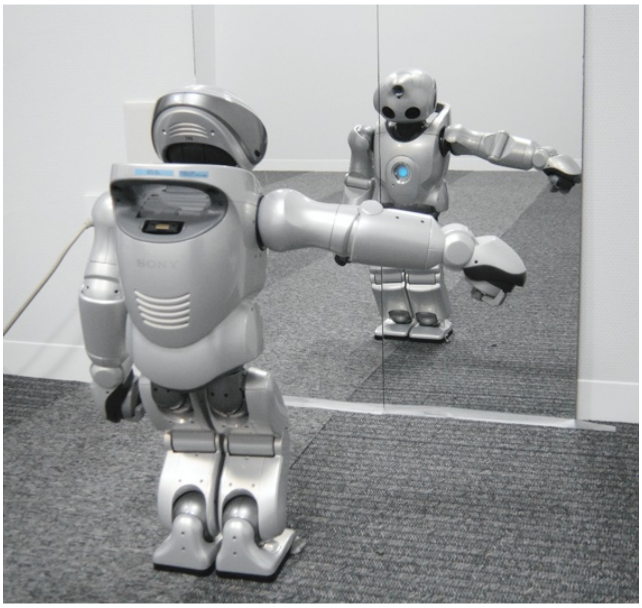

Figure 2. *A humanoid robot stands before a mirror and performs various motor behaviors thus observing what visual body-images these behaviors generate.*

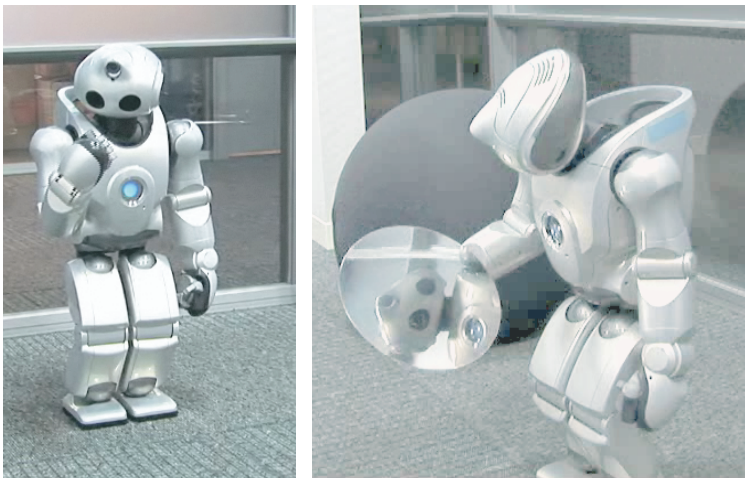

Figure 3. *The robot can also inspect its own body parts directly (left) or use a smaller mirror to inspect areas that are difficult to see otherwise, such as its own head (right).*

types (i.e. posture image schemata), posture nodes, motor behaviors and posture names. The network is similar to the one used in the Grounded Naming Game experiments discussed in an earlier chapter (see Figure 7 in Steels & Loetzsch, 2012) with postures nodes instead of nodes for individuals and with additional nodes for the motor behaviors that can achieve postures. Each link is scored and the link with the highest score is preferred when the network is used:

- To request a posture, the speaker chooses the name associated to this posture with the highest score in his private memory and uses that name.
- To interpret a name, the listener chooses the highest scoring association between this name and the posture in his private memory, then chooses the highest scoring behavior that is associated with this posture and performs the associated behavior.
- To check whether the game was successful, the speaker matches the visual prototypes of postures with the perception of the posture taken on by the hearer, and retrieves from there the highest scoring association to a posture node. If this is the one originally chosen, the game is a success.

Note that the same posture can be associated with many different names (synonymy) and the same name with several postures (meaning uncertainty).

Here is a more detailed description of each of the nodes in this semiotic network (see Figure 4 for a schematic depiction):

1. *Sensory experience*: The vision system not only segments the robot body against the background (Figure 5) but also extracts a series of features that characterize the shape of the robot body. Because these features have to be translation and scale invariant, we use normalized central moments (Mukundan, 1998; Hu, 1962), as defined in more detail in (Steels & Spranger, 2009). Values for each of these features are combined into a feature vector that constitutes the sensory experience of a sensed visual body image at a particular moment in time. When the robots are watching another robot, they perform perspective reversal on the visual image, which means that they perform a geometric transformation depending on their relative position with respect to the other robot.

2. *Prototypes*: The specific visual feature vectors must then be classified in terms of prototypes of body postures. Prototypes are condensed statistical representations of past experience. They are computed by averaging all feature values of the feature vectors for all instances. The best matching prototype given a particular sensory experience is the nearest neighbor based on Euclidean distance (see a more

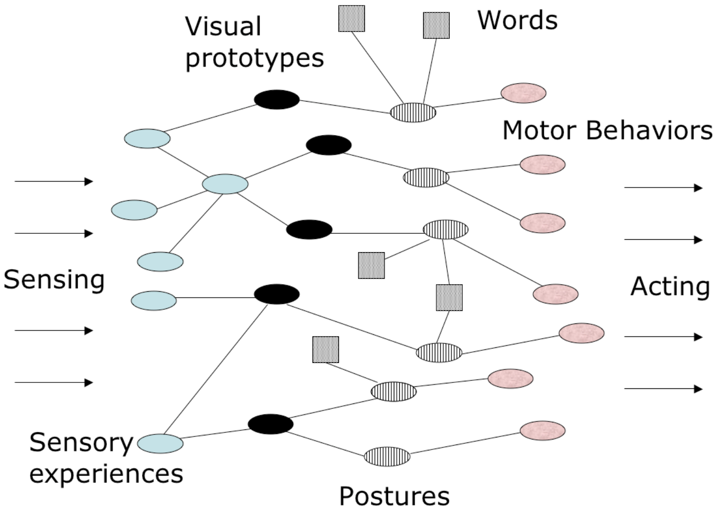

Figure 4. *Semiotic network linking sensory experiences, visual prototypes of postures, nodes for postures acting as mirror neurons, and nodes triggering the motor behavior that achieves the posture. Nodes for words (shown as squares) are associated with the posture nodes.*

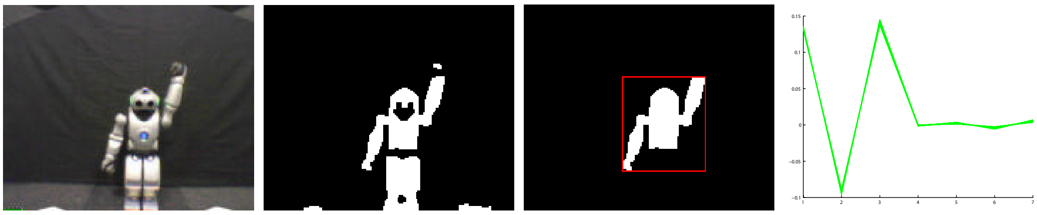

Figure 5. *Aspects of visual processing. From left to right we see the source image, the foreground/background separation, the result of object segmentation (focusing on the upper torso only), and the feature signature of this posture(explained in the text).*

formal definition in the appendix of Steels & Spranger, 2009 and Spranger et al., 2009 for a similar unsupervised approach). An example of a prototype is shown in Figure 5 (right most). The seven features are displayed on the x-axis and their values (scaled) on the y-axis. The graph shows a band connecting minimum and maximum boundaries of each feature values for this prototype (which is very thin in this particular case, implying that the feature values have to precisely match).

3. *Postures*: Each visual prototype is linked to a posture node and the posture node itself is linked to a motor behavior to achieve the posture. These posture nodes therefore function like *mirror neurons*, in the sense that they become active both when the robot sees the visual image of a posture by another robot (possibly himself) that matches best with the prototype of this posture AND when the robot is about to perform the motor behavior that generates this posture.

4. *Motor Behavior*: The motor behavior node is a control node that is connected to the complex commands and sensory feedback loops needed to achieve the posture.

2.2. Diagnostics, Repairs and Alignment

We now examine the strategies that agents can use to build up and maintain these semiotic networks. These are similar to the ones already discussed in other chapters for the Grounded Naming Game (Steels & Loetzsch, 2012), the Color Language Game (Bleys, 2012) or the Spatial Language Game (Spranger, 2012).

The scores of the link between a posture node and its motor control program is in the present experiment acquired through *kinesthetic teaching* and stored in long term memory. This means that the experimenter decides how many postures need to be learned and the robot then acquires what motor control program should be associated with this posture with full certainty. All robots share the same set of posture nodes and motor control programs. This strong assumption is removed in the next section.

The score of the link between a sensory experience and a posture prototype is determined dynamically in relation to the present context. The image is segmented so that the robot body appears against the background (see Figure 5, left), the torso and arms are segmented out, and the features characterizing this shape, i.e. the normalized central moments, are computed (Figure 5, far right). This feature vector is compared with the feature vectors that define the posture prototypes using a Euclidean distance measure. The score is equal to how well each prototype fits (see details in Steels & Spranger, 2009).

How are the links between the visual prototypes of a posture and the node for each posture acquired? This is where motor babbling before the mirror plays a crucial role because it generates the conditions under which learning can take place: The robot stands before the mirror, selects a posture, and activates the corresponding motor behavior. This motor behavior generates a sensory image. The image undergoes a 180 degree geometric transform to take into account that the robot looks at a mirror image of itself and then a feature vector is computed. Subsequently, agents simply store the obtained feature vector and re-estimate the prototypical postures given all previously encountered samples, so that the prototype better reflects the visual image schemata of this posture.

Once a set of posture nodes and their associated visual prototypes and motor programs are established, the development of a shared vocabulary can use the same strategy as for the Naming Game dynamics already introduced in an earlier chapter Steels & Loetzsch (2012), based on the following repairs:

1. *Speaker has no name*:
   - Diagnostic: The speaker finds no name associated with the posture that he wants the hearer to perform.
   - Repair: The speaker introduces a new name in the form of a random combination of syllables, and creates an association between the posture and the name in long term memory with an initial score σ_{init}
2. *Hearer does not know the name*:
   - Diagnostic: The hearer encounters a new name that is not yet stored in his vocabulary.
   - Repair: The hearer signals failure and infers a possible posture based on the example shown by the speaker. He stores the association between the name and the posture with an initial score σ_{init}. It is possible that there was no posture yet matching with the perception of the the example shown by the hearer in which case a new prototype node is added to memory.
3. *Hearer uses name differently*:
   - Diagnostic: The action taken by the hearer did not yield the body posture that the speaker expected.
   - Repair: The speaker signals failure and the hearer infers the posture based on the example shown by the speaker. The hearer then stores the association between the name and the posture in his semiotic network, if it does not exist yet with an initial score σ_{init}, or else he increases the score of this association.

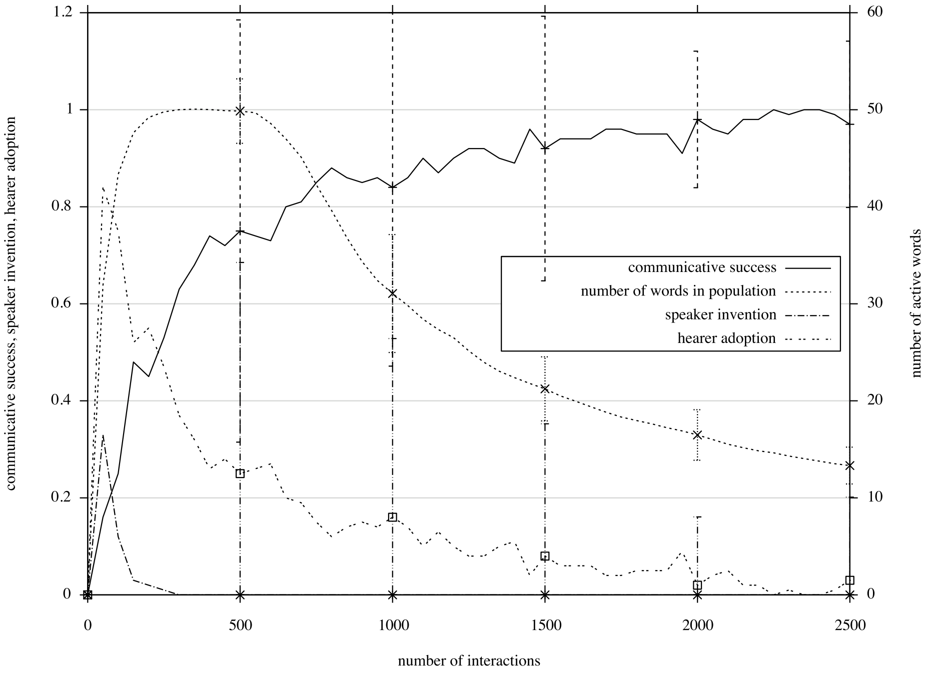

Figure 6. *This figure shows a series of 2500 games in a population of 10 agents naming 10 postures. Robots have first coordinated their visual body-images and motor behaviors by standing before a mirror and observing their visual appearance in relation to certain motor behaviors. The number of language games (in this case 2500 games) are shown on the x-axis, each game involving two agents. The running average of communicative success as well as invention and adoption frequency is shown (left y-axis) and the average vocabulary size (right y-axis).*

There is also the necessary alignment using a lateral inhibition strategy familiar from the Naming Game:

- Alignment after a successful game: Both speaker and hearer increase the score of the used association with $\delta_{success}$ and diminish competing associations with $\delta_{inhibit}$.
- Alignment after a failed game: Both speaker and hearer decrease the score of the used association with δ_{fail}.

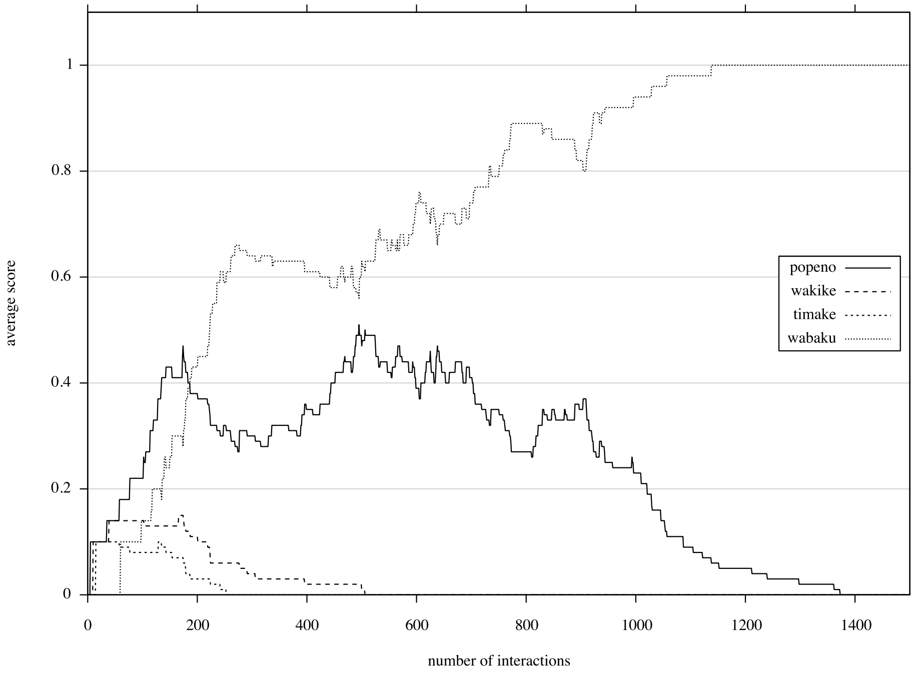

Figure 7. *This graph shows the average score for all the different words in the lexicon of all agents which are competing for naming the same posture. The words are "popene", "wakike", "timaki", and "wabaku". The winner-take-all dynamics that coordinates the vocabulary among the agents is clearly visible. "Wabaku" wins the competition.*

2.3. Experimental Results

When a population of agents uses this language strategy, we get the results as shown in Figure 6. The graphs plot for the first 2500 posture games the global behavior of the population *after* each individual has coordinated motor behaviors and visual body-images through the mirror. The y-axis plots the communicative success, vocabulary-size, and invention and adoption frequency. Figure 7 shows the average score in the population for the different names competing for the same posture. We see clearly that a winner-take-all situation arises after about 300 games, with one word dominating for naming this particular posture.

3. Coordination without mirrors

We have seen that once coherent links exist between the image schema for a posture and the motor behavior that generates this posture (here mediated by the posture nodes acting as mirror neurons), it is straightforward for a group of agents to self-organize a vocabulary of names which can be used as commands to ask another robot to achieve it, and thus for playing the Posture Game. Now we turn to the question whether robots could also coordinate visual body-image and motor behaviors through language and *without* using first a mirror. It turns out that the answer is positive, showing the potential of language to coordinate perception and behavior among autonomous agents without any form of telepathy, central control or prior design. We do not claim that language is the only way in which autonomous agents can coordinate motor behaviors and visual body images, indeed the previous section already has shown that inspecting your body while performing a movement is very effective, and the next chapter will show that if agents have access to a simulation they can even better guess the meaning of unknown posture names and learn which visual posture prototypes get linked to which motor behaviors.

Because the experimenter no longer determines the set of possible postures or shows through kinesthetic teaching how they should be achieved, the networks built up by the agents no longer contain nodes for postures. Instead we use networks as shown in Figure 8. Nodes for words are now linked with the visual prototypes of postures on the one hand and with motor behaviors on the other. It is the words that act as the glue between the two.

The script for the Posture Game in this experiment is the same as the one used earlier, except that the speaker now chooses a visual prototype (because there are no longer explicit posture nodes) and motor behaviors are associated with names:

1. The speaker chooses randomly a visual prototype of a posture from his semiotic network.

2. The speaker retrieves the name associated with this prototype from his private vocabulary and transmits that to the hearer.

3. The hearer retrieves the motor behavior associated with the highest score to this name in his own vocabulary and evokes the motor control program associated with it.

4. The speaker categorizes the visual image generated by the hearer's movements in terms of the prototypical postures in his own semiotic network and

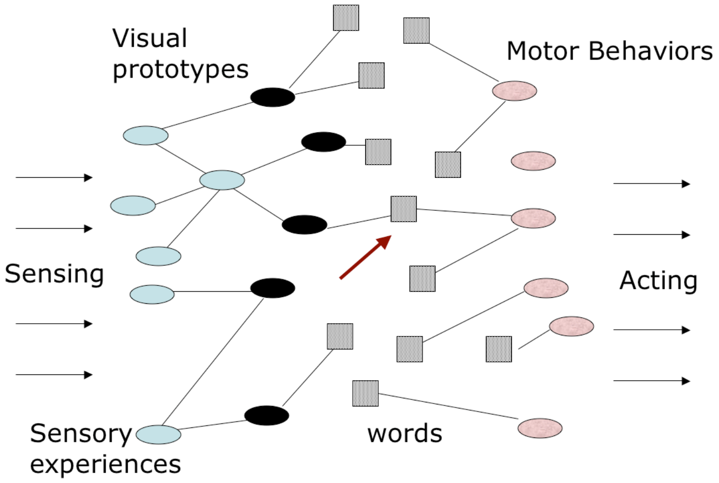

Figure 8. *Semiotic networks built by agents in the experiment without the mirror. There are no nodes for postures linking visual prototypes with motor behaviors. Nodes for words (indicated by squares, as pointed at by the arrow) are either linked to posture prototypes or to motor behaviors or both.*

checks whether this prototype is associated to the name of the posture prototype he originally chose in step 1.

5. If this is not the case, the speaker signals failure, otherwise success.

3.1. Diagnostics, Repairs and Alignment

The strategies for invention, adoption, and alignment have to be changed as well.

1. *Speaker has no name*:

   - Diagnostic: There is no name associated with a visual prototype in the speaker's semiotic network (step 2 fails).
   - Repair: The speaker invents a new name, associates it with the prototype, and uses the name.

   This repair has the effect that a new name for a posture prototype is introduced but the connection to a possible motor program remains unknown.

2. *Hearer does not know the name*:
    - Diagnostic: The hearer signals failure.
    - Repair: If the speaker has a motor program associated with the name, it is executed. If the hearer has a visual prototype for the sensory image created by the speaker, he stores a link between this prototype and the word or increases the score of the link if it exists already.
3. *Hearer uses name differently*:
    - Diagnostic: The action taken by the hearer did not yield the body posture that the speaker expected.
    - Repair: The speaker signals failure and performs the action. The hearer infers the posture and stores the association between the name and the posture in his semiotic network, if it does not exist yet, or else he increases the score of this association.

Note that agents change the connections between words and visual prototypes only when they are speaker and they change the connections between words and motor behaviors only when they are hearer. Alignment is the same as in the mirror experiment.

3.2. Experimental Results

Results of this second experiment for a population of 5 agents and 5 postures are shown in Figure 9. The graphs show the global behavior of the population focusing on the first 5000 language games. 100 % success is reached after about 5000 games and stays stable (unless new postures get introduced and then the networks of all agents will expand to cope with this). Already after 3000 games there is more than 90 % communicative success. The graph shows the typical overshoot of the lexicon in the early stage as new words are invented and a phase of alignment as the agents converge on an optimal lexicon, which is now 5 words to name each of the postures. The frequency of invention and adoption is also shown to die down as the lexicon stabilises. So this is already a very important result. Clearly agents are able to self-organize a lexicon of commands even if no prior lexicon exists and even if they are not pre-programmed nor have been able to learn mappings from motor behaviors to visual postures before lexical self-organization starts.

Communicative success does not necessarily mean that agents relate all visual prototypes 'correctly' to the corresponding motor behaviors. Nevertheless, Figure

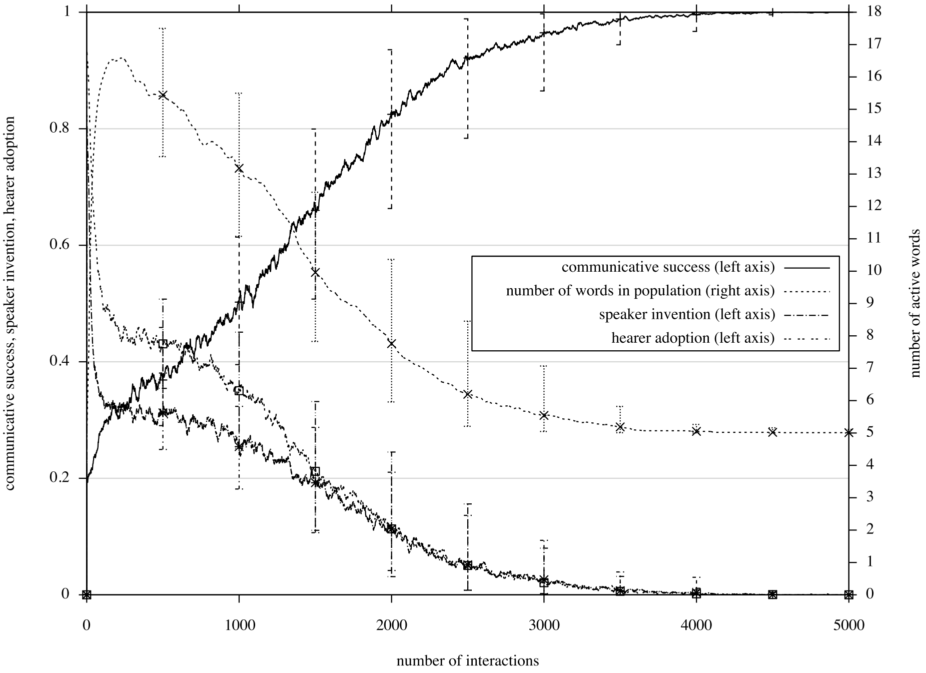

Figure 9. *Results of the Posture Game played by a population of 5 agents for 5 postures. This time, robots have not coordinated their visual body-images and motor behaviors by using a mirror but only through language. The x-axis plots the number of language games. The running average of communicative success and average lexicon size as well as invention and adoption frequency are shown on the y-axis.*

10 shows that only correct mappings are established. The figure plots the aggregated strength (black - strong mapping, white - no mapping) of the relation between visual prototypes (x-axis) and motor behaviors (y-axis) as mediated by language. Aggregated results for 100 repetitions of the same experimental trial are shown in the diagram. The diagram shows agreement among the agents (computed by taking the product of the strength of visual prototype to motor behavior mappings over all agents in all experiments). If there is a single column between a particular visual prototype p_i and its corresponding motor behavior m_i with strength 1.0 (black), then this means that all agents agree on it. This is the case here for all prototypes and

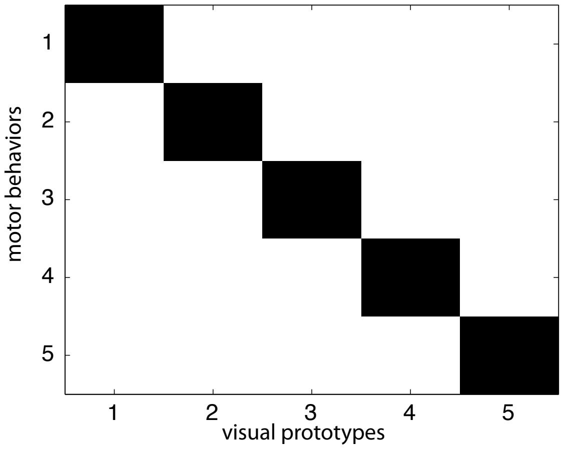

Figure 10. *Relation between visual prototypes and motor behaviors for a population of 5 agents after they have reached a steady state. The diagrams combine the results of 100 different experiments, showing the aggregated strength (black - strong, white - none) of the relation between visual prototypes and motor behaviors as mediated by language. The diagram shows that only correct links (e.g., motor behavior 1 is only linked to visual prototype 1) are present and that all agents agree on the correct mapping.*

motor behaviors in all 100 experiments. In other words, total coherence with respect to visuo-motor coordination has been reached by all agents in all experiments.

This result is remarkable because there appears to be a chicken and egg situation. If all agents have a correct association, then a lexicon can easily self-organize as shown in the mirror experiment. And alternatively if there is a shared lexicon (for example pre-programmed), then we can imagine the agents to learn a new association between visual prototype and motor behavior easily. But how is it possible that agents coordinate their lexicons AND acquire coherent associations between visual body-image of other and motor behavior of self, without any form of telepathic exchange or prior design, and without any nodes functioning as mirror neurons linking the two?

Moreover, an interesting phenomenon can be further observed, namely that two agents can arive at a successful communication, but with rather different internal semiotic networks, in the sense that each agent associates a different visual body

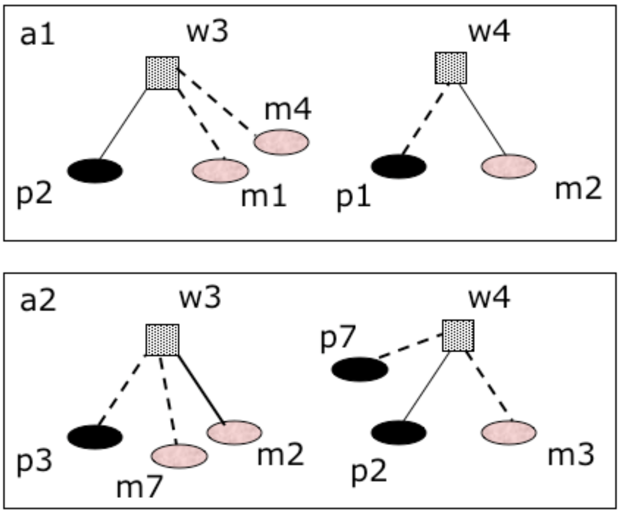

Figure 11. *Successful games may arise between two agents even without visuo-motor coordination. This figure shows an intermediate state of two agents where different prototypes or motor behaviors are associated with the words. In an ideal situation motor behavior w_i is associated with m_i and with prototype p_i. The word w_3 can be successfully used by a_1 to evoke m_2 (giving rise to p_2) by a_2.*

image with a motor program (see Figure 11). For example, agent a_1 may have associated the visual prototype p_2 with the word w_3 and motor-behavior m_2 with the word w_4. Another agent a_2 may have associated w_4 with p_2 and w_3 with m_2. The agents have in addition other associations but with a lower score. There is no visuo-motor coordination within each agent, assuming for example that p_2 is the prototype corresponding with m_2, but nevertheless agents can successfully communicate. When a_1 wants to see p_2 he communicates w_3, and a_2 will perform m_2 which indeed reproduces p_2 for a_1. The same with turns taken is true for w_4, p_2 and m_2. So we see that success in the language game does not necessarily mean that each agent uses the same word both for recognising and for producing a particular posture. They can have perfect communicative success without visuo-motor coordination.

The situation changes completely however when a third agent a_3 gets involved. He will be confronted with two competing words for achieving p_2: w_3 from a_3

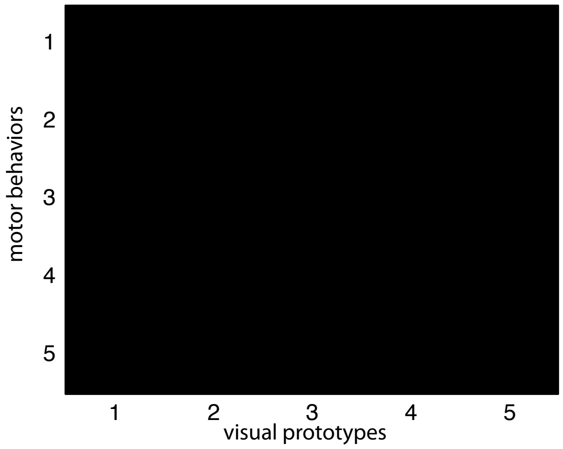

Figure 12. *This shows the mappings agents establish in populations of 2 agents after reaching a stable state (results of 100 experimental trials). All combinations of visual prototype to motor behavior mappings occur (all mappings are black). Populations of 2 agents are, therefore, not able to agree on the correct mappings from the visual space to the motor space. Compare this to Figure 10 which shows a more ordered case where only correct mappings survive.*

and w_4 from a_2. Only one of them can survive the lateral inhibition dynamics, pushing also the other agents to choose one or the other word for p_2. So three or more agents can only communicate successfully when prototypes and motor-behaviors *are* coordinated within each agent. Given this observation, we predict that in a population with only two agents, no visuo-motor coherence will arise, despite successful communication. Experiments show that this is indeed the case (see Figure 12).

4. Improving meaning guesses through self-simulation

We now address the question whether it is possible to improve the efficiency of the overall system. When the speaker does not know which motor behavior corresponds to a posture he would like to see achieved, he has to make a guess (step 2) and when the hearer encounters an unknown word, he has to choose a motor behavior that could be associated with the visual image he is perceiving (step 7). In

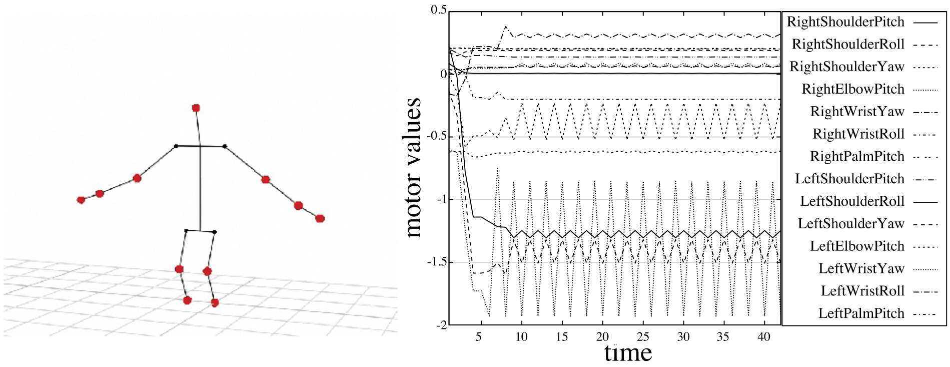

Figure 13. *Paired schematic kinematics model (left), control stream (right) while the robot executes a waving action.*

the interaction pattern used in the previous experiment, these choices are made randomly, which implies that the choice is correct in only $\frac{1}{A}$ of the cases, with A equal to the number of possible actions, which explains worsening performance when the number of actions increases. We now show how this choice can be improved significantly by using a simulation-based approach (Barsalou, 1999; Feldman & Narayanan, 2004; Feldman, 2006; Steels & Spranger, 2008a).

The basic idea is as follows. Robots feature already a predictive model of their own body movements which they use for adaptive control. This model is based on a model of the robot's own physical body (including information about body parts, their size and shape, how they are attached to each other) and a schematic kinematics model and it integrates ongoing motor commands and proprioceptive motor streams (see Figure 13). Using this model, the robot can predict the proprioceptive effect of a command and thus monitor whether it is appropriately executed. This motor body model can also be decoupled from ongoing behavior so that the robot can simulate a complete action.

Second, a simulated visual body image can be generated from the simulated motor body image. For example, the simulated motor body model contains information about the angles of the different joints which can be used to internally visualize the position of the joints with respect to the body. This visual 'imagining' is from the robot's own frame of reference. In order to create an expectation of what the same behavior performed by another robot would look like, a third step is needed: The robot has to perform a perspective reversal by performing a geometric transform on this visual image, based on knowing the position of the other robot. Finally, similar visual processing and categorisation can be carried out on this simulated visual body image as on the 'real' experienced visual body image of the other robot, specifically centralized moments can be computed again to extract the features needed for gesture categorisation.

Given this competence in visual imagination, robots can now improve drastically the quality of their guesses in steps 2 and 7 of the interaction pattern. Speakers can use the simulated visual body image of the other robot to make a better guess of the correct motor-behavior given a visual prototype of the posture they want to see adopted by the other robot. This is done by a process that starts from a random choice of action from the action repertoire, simulates that choice to construct a visual body image of the other robot, and compares this to the desired posture. In case of a failure, a new action is chosen until a reasonable match is found. The hearer goes through a similar procedure if he has to guess the meaning of an unknown word. As mentioned earlier, simulated behavior will always deviate from actual behavior and so this process does still not give an absolutely certain guess.

Figure 14 shows the important impact on performance of this simulation-based approach. The base-line case (marked 0.0) is the one used in earlier experiments, i.e. where the agent has a $\frac{1}{A}$ chance to choose the correct action. It shows the slowest rise towards communicative success. The best case (marked 1.0) is one where the agents manage to always guess correctly both as speaker and as hearer what motor behavior achieves a visual prototype of a posture based on self-simulation, imagination, perspective reversal, and visual categorisation. This case approaches the performance in the mirror experiment where agents first learned the mapping between visual body image and motor body image by standing in front of a mirror. In the intermediary cases we see that that the higher the probably of correct guessing the faster the population reaches full communicative success.

Figure 15 shows the semiotic dynamics for the vocabulary of the agents for the same series of experiments. There is always a phase of invention and spreading with a typical overshoot in the size of the vocabulary, which then settles down on an optimal vocabulary as agents align their word meanings. We see that the overshoot of words is much smaller when the quality of guessing improves, which implies that time to convergence is also much shorter.

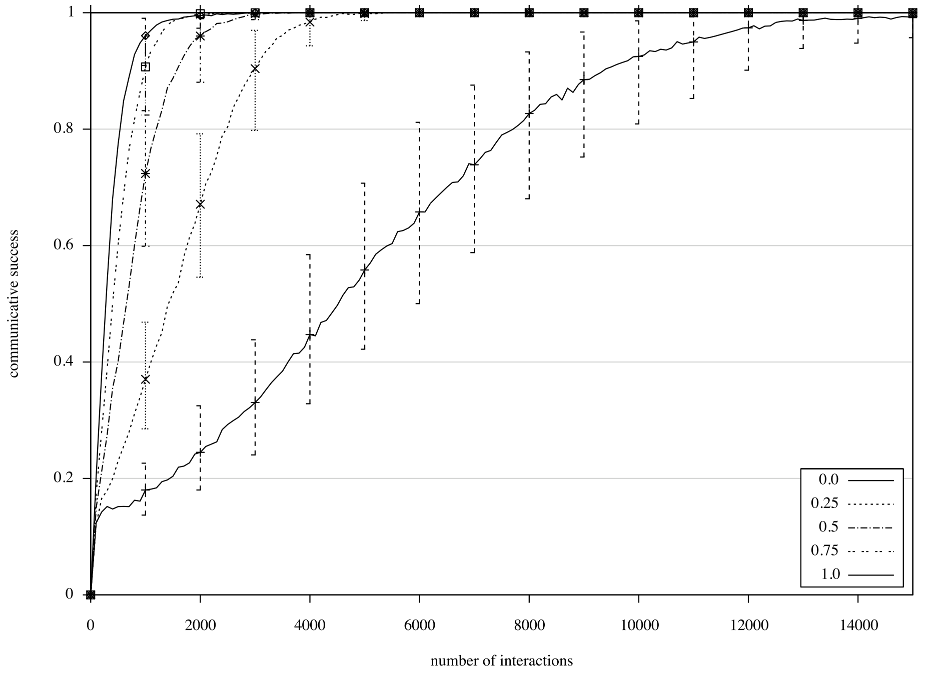

Figure 14. *Influence of simulation on the performance of the agents in the Posture Game (5 agents, 10 actions). The figure shows the communicative success averaged over 100 different experimental runs, each for a single parameter.* 0.0 *is the case were the speaker guesses the correct motor-behavior with the base line probability* $\frac{1}{10}$. 1.0 *means that speaker and hearer "guess" correctly every time they have to choose a motor-command to match with a desired visual prototype (as speaker) or interpret an unknown word (as hearer).*

5. Conclusions

This chapter tackled the emergence of a communication system by which agents can command each other to take on certain bodily postures. This challenge required not only that a vocabulary gets established, but also that agents acquire a mirror system relating visual prototypes of postures with behavioral programs to achieve them, indeed the latter is even a bigger challenge than the emergence of a vocabulary. Three experiments were performed: one where agents bootstrap a mirror system by inspecting themselves before a mirror, a second experiment where agents use language to coordinate visual and motor images, and a third experiment which

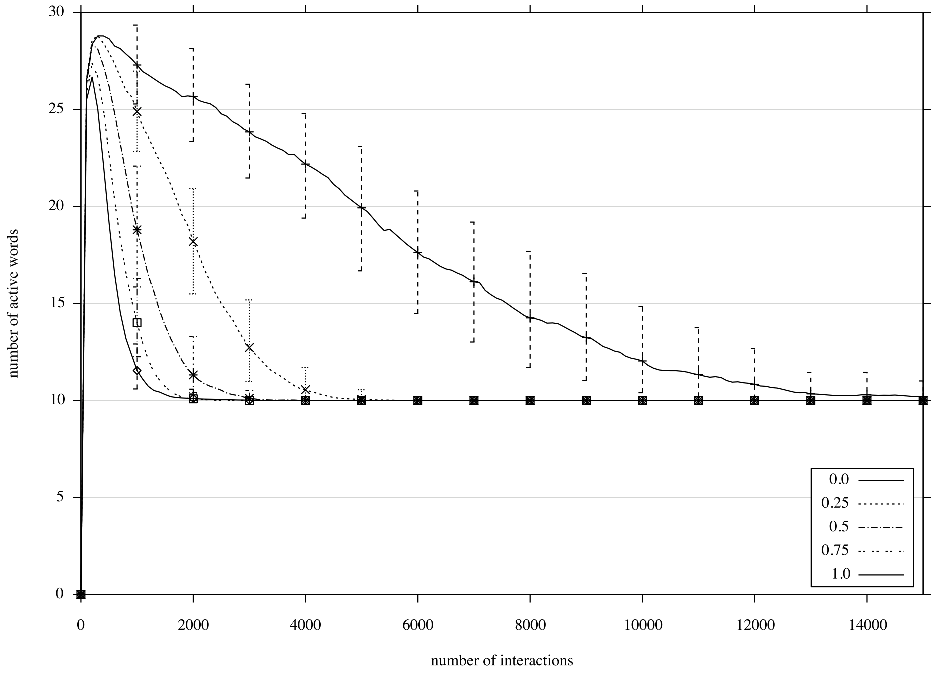

Figure 15. *Study of the influence of simulation on the performance of the agents in the game (5 agents, 10 actions). The graph depicts the number of words (average lexicon size across all agents) averaged over 100 experiments for a given parameter.* 0.0 *is the case where the speaker guesses the correct motor-behavior with the base line probability* $\frac{1}{10}$. 1.0 *means that the speaker "guesses" correct every time he guesses a motor-command.*

exploits the added value of a simulation in guessing the meaning of unknown postures.

The three experiments together argue for three important points. First, grounded representations linking the different motor, visual and proprioceptive domain can greatly enhance the success of agents in establishing communication systems (Section 2). However, even if these mappings are not a priori established other agents can serve as a mirror system and enable the population as a whole to find the right mapping between different sensorimotor spaces (Section 3). Lastly, the impact of reliable mappings between different sensorimotor spaces can be quantified and simulation capabilities (Section 4) allow to interpolate between the a priori established mapping condition (Section 2) and the no mapping condition (Section 3).

This chapter illustrates how a whole-systems approach is able to deal with fundamental problems in cognitive science that remain mysterious otherwise. It is only by the full integration of all aspects of language with sophisticated sensory-motor intelligence that agents were able to arrive at a shared communication system that is adequate for the game. The chapter also illustrates how the theory of linguistic selection operates. Although the agents rely as much as possible on their cognitive capacities (such as predictive motor control) to come up with a communication system that has the required expressive power and social conformity needed to reach persistent communicative success, they still have to make guesses and create variants that then will be selected for in the collective semiotic dynamics.

Acknowledgements

This research has been carried out at the Sony Computer Science Laboratory in Paris with partial support from the ECAgents project, funded by the EU Future and Emerging Technologies program (IST-FET) as IST-1940, and the ALEAR project, funded by the EU Cognitive Systems program. We are extremely grateful to Masahiro Fujita, Hideki Shimomura and their team at the Sony Intelligent Systems Research Lab in Tokyo for giving the opportunity and support to perform the robotic experiments, and to Martin Loetzsch for his help in setting up and carrying out the experiments.

References

Barsalou, Lawrence (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.

Billard, Aude (2002). Imitation: a means to enhance learning of a synthetic protolanguage in an autonomous robot. In Kerstin Dautenhahn, Christopher Nehaniv (Eds.), *Imitation in Animals and Artifacts*. Cambridge Ma: The MIT Press.

Blanke, Olaf, Zenon Castillo (2007). Clinical neuroimaging in epileptic patients with autoscopic hallucinations and out-of-body experiences.case report and review of the literature. *Epileptologie*, 24, 90–96.

Bleys, Joris (2012). Language strategies for color. In Luc Steels (Ed.), *Experiments in Cultural Language Evolution*. Amsterdam: John Benjamins.

Demiris, Yiannis, Matthew Johnson (2003). Distributed, predictive perception of actions: a biologically inspired robotics architecture for imitation and learning. *Connection Science*, 15(4), 231–243.

Feldman, Jerom, Srini Narayanan (2004). Embodied meaning in a neural theory of language. *Brain and Language*, 89, 385–392.

Feldman, Jerome (2006). *From Molecule to Metaphor: A Neural Theory of Language*. Cambridge, MA: Bradford Books.

Fujita, Masahiro, Yoshihiro Kuroki, Tatsuzo Ishida, Toshi Doi (2003). Autonomous behavior control architecture of entertainment humanoid robot SDR-4X. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 960–967.

Gallese, Vittorio, Christian Keysers, Giacomo Rizzolatti (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, 8(9), 396 – 403.

Gallese, Vittorio, George Lakoff (2005). The brain's concepts: the role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3/4), 455–479.

Ghahramani, Zoubin, David Wolpert (1997). Modular decomposition in visuomotor learning. *Nature*, 386(6623), 392–395.

Hoffmann, Matej, Hugo Marques, Alejandro Arieta, Hidenobu Sumioka, Max Lungarella, Rolf Pfeifer (2010). Body schema in robotics: A review. *IEEE Transactions on Autonomous Mental Development*, 2(4), 304–324.

Hu, Weiming (1962). Visual pattern recognition by moment invariants. *IEEE Transactions in Information Theory*, 8, 179–187.

Mukundan, Ramakrishnan (1998). *Moment Functions in Image Analysis: Theory and Applications.* New York: World Scientific.

Pfeifer, Rolf, Joe Bongard (2006). *How the Body Shapes the Way We Think.* Cambridge, MA: The MIT Press.

Pulvermüller, Friedemann (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, 6(7), 576–582.

Ramachandran, Villayanur, William Hirstein (1998). The perception of phantom limbs. *Brain*, 121, 1603–1630.

Rizzolatti, Giaccomo, Leonardo Fogassi, Vittorio Gallese (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature Reviews Neuroscience*, 2(9), 661–670.

Rizzolatti, Giacommo, Michael Arbib (1998). Language within our grasp. *Trends in Neurosciences*, 21(5).

Rizzolatti, Giacomo (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology*, 210(5), 419–421.

Rosenfield, Israel (1988). *The Invention of Memory: A New View of the Brain.* New York: Basic Books.

Spranger, Michael (2012). The co-evolution of basic spatial terms and categories. In Luc Steels (Ed.), *Experiments in Cultural Language Evolution*. Amsterdam: John Benjamins.

Spranger, Michael, Sebastian Hoefer, Manfred Hild (2009). Biologically inspired posture recognition and posture change detection for humanoid robots. In *Proceedings of ROBIO'09: IEEE International Conference on Robotics and Biomimetics.* IEEE.

Spranger, Michael, Martin Loetzsch (2009). The semantics of sit, stand, and lie embodied in robots. In N. A. Taatgen, H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society (Cogsci09)*, 2546–2552. Austin, TX: Cognitive Science Society.

Steels, Luc, Martin Loetzsch (2012). The grounded naming game. In L. Steels (Ed.), *Experiments in Cultural Language Evolution*. Amsterdam: John Benjamins Co.

Steels, Luc, Michael Spranger (2008a). Can body language shape body image? In Seth Bullock, Jason Noble, Richard Watson, Mark Bedau (Eds.), *Artificial Life XI: Proceedings of the Eleventh International Conference on the Simulation and Synthesis of Living Systems*, 577–584. The MIT Press.

Steels, Luc, Michael Spranger (2008b). The robot in the mirror. *Connection Science*, 20(2-3), 337–358.

Steels, Luc, Michael Spranger (2009). How experience of the body shapes language about space. In Hiroaki Kitano (Ed.), *IJCAI'09: Proceedings of the 21st international joint conference on Artifical intelligence*, 14–19. San Francisco: Morgan Kaufmann.

Talmy, Leonard (2000). *Toward a Cognitive Semantics, Typology and Process in Concept Structuring*, vol. 2. Cambridge, Mass: MIT Press.