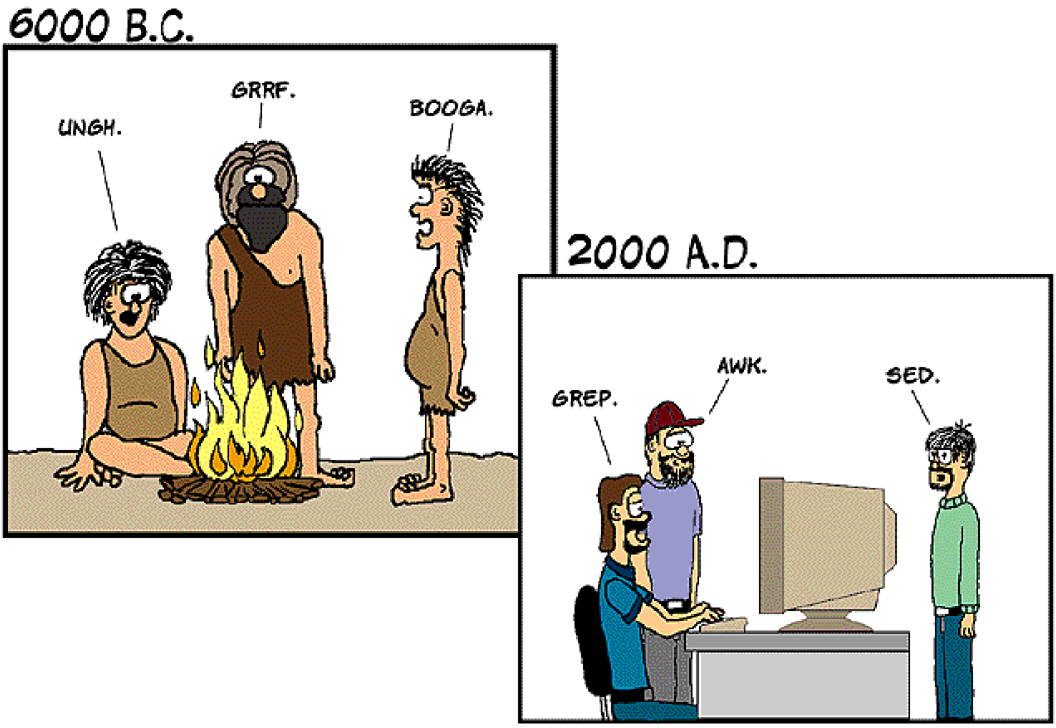
Joint Attention, Language Evolution, and Development



Johan Kwisthout

Master Thesis in Artificial Intelligence

Joint Attention, Language Evolution, and Development

By Johan Kwisthout, June 2006

Master Thesis in Artificial Intelligence

Radboud University Nijmegen

Faculty of Social Sciences

Student number 0234265

In partial fulfilment of the requirements for the degree of
Master of Science in Artificial Intelligence

Daily supervisors:

Dr. A.F.J. Dijkstra

Dr. W.F.H. Haselager

Manuscript committee:

Dr. A.F.J. Dijkstra

Dr. W.F.H. Haselager

Dr. P. Vogt

Table of Contents

Abstract **5**
Acknowledgements **5**
1 Introduction **6**
1.1 Joint attention and Theory of Mind 7
1.2 Joint attention, language development, and Theory of Mind 10
1.3 Joint attention and language evolution 11
1.4 Investigating language evolution 13
1.5 Using language games as evidence for language evolution 14
1.6 Research questions addressed in this thesis 15
1.7 Subquestions 16
2 Model, simulation, and empirical research **17**
2.1 Language Games 17
2.2 Language games versus language learning 20
2.3 Reported experimental results 20
2.4 Implementation of joint attention in the observational game 21
2.5 Simulation settings 22
2.5.1 Method 22
2.5.2 Coherence rate 23
2.5.3 Coherence development over time 24
3 Results **26**
3.1 Preliminaries 26
3.2 Language game results 27
3.2.1 Guessing Game 28
3.2.2 Selfish Game 28
3.2.3 Observational Game 29
3.2.4 Joint Attention enhancements 29
3.3 Using three attributes 29
3.4 Using three agents 30
3.5 Using ambiguous environments 31
3.6 Iterated learning game with fixed labels and fixed roles 33
3.7 Discussion of these results 34
4 Discussion **36**
4.1 General discussion 36
4.2 Language games as model for language evolution 38
4.3 Conclusions 38
4.4 Further research 40
References **41**
Appendices **44**
1 All results of language games 44
2 Pseudocode of the language games 46
2.1 guessing game 46
2.2 observational game 48
3. Settings and parameters compared with De Vogt and Coumans (2003) 50
4. Construction of the environments 50
5 Protocols of the False Belief and Opaque Context tests 51
5.1 False Belief Test 51
5.2 Opaque Context Test 51

Abstract

In this thesis I explore the relation between joint attention, language evolution and development using so-called *language games*, computational simulations of a population of agents that try to develop a grounded lexicon using communicative actions in a particular environment. Based on previous research that involved language games such as the *guessing* and *observational* game, where feedback respectively shared attention were the mechanisms that facilitated lexical grounding, I developed an enhanced observational game that used following and directing attention as primary mechanisms. Using these more advanced stages of joint attentional development, I investigated whether evidence could be found for the *escalator model*, in which joint attention is an evolutionary precursor for both Theory of Mind and language, which build on each other to further develop. The results of the simulations suggest that this model is too simple, in the sense that joint attention *alone* is not the 'crucial ingredient' for the evolution of language and Theory of Mind.

Acknowledgements

I owe very much to my daily supervisors Ton Dijkstra and Pim Haselager, who agreed to guide this thesis research from my initial vague ideas to this finished thesis. Their ideas, suggestions and critiques greatly helped to focus and improve my research and this thesis. Especially after I obtained my masters degree in computer science, and embarked on a PhD project, they encouraged me to find – or make – time to complete this thesis. Apart from supervising, they also took time to introduce me to colleagues and gave me the opportunity to present my work – and comment on these presentations. I'd like to thank Ton for providing me with funding as a research assistant, which was an important support: Not only financially in an 'in-between jobs' period, but also as a valuable research experience.

Furthermore, I want to thank Paul Vogt for his comments on my thesis proposal and for his willingness to make time to be a member of the manuscript committee.

Last, but certainly not least, a big 'thank you' to Judith, for love, caring, and drawing the beautiful picture on page 5.

1 Introduction

Anne and her toddler are sitting on the couch together, watching pictures in a booklet. The pictures show animals of all sorts. “Look, Timmy!” Anne says, and points to the cow on one of the pages, “Look at this picture here. What sound does this animal make?” “Moo!” Timmy answers. “Yes, that’s right; a cow says ‘moo’. Let’s take a look at the next page. Oh, that’s a pussy-cat, just like Tiger! Do you see Tiger over there, in her basket? This cat looks like Tiger, don’t you think?” Timmy looks where Anne is pointing, sees the cat, and nods.

Perhaps the same location, but one hundred thousand years earlier. A small group of men from the local tribe is hunting a mammoth. Until now, their attacks have had little impact. The mammoth runs to the open ground, and the hunters fear that they might lose it out there. They have not eaten properly for quite some time and the tribe is near to starvation. Suddenly, Moonwatcher notices a small niche in the nearby mountain hill. If they manage to direct the mammoth to that niche, they can trap it there and kill the mammoth. He calls on Old One and points to the niche, then points at Old One and then to the opposing side of the mammoth, making a ‘pushing’ gesture in the direction of the niche. Old One immediately understands they have to attack the mammoth from both sides in order to force it to run to the niche.

What connects these seemingly diverse situations? Crucial in both is the participants’ ability to engage in *joint attention* in order to understand each other. For young children, the ability to share attention with an adult with respect to a third object or actor is a very important step in their language development. In evolution, the ability to engage in joint attention may have been the crucial *mechanism* that enabled mankind to rise from stone age to modern culture or technology in relatively short time, as suggested by Tomasello (1999), making *cultural learning* possible. To be able to engage in joint attention might also be a crucial prerequisite in language evolution.

Simulating *lexical grounding* with *language games* is one possible method that has been used to investigate the role of joint attention in language development and evolution. In these language games, computer agents (either robots or software agents) play sequences of games that involve naming and referencing objects, thus trying to establish a common association between arbitrary labels to objects or attributes of objects. Although robot language games, e.g., the Talking Heads experiments of Luc Steels (Steels, 1999a), use joint attention as a mechanism to allow the robots to know the topic of the language game (i.e., the object that is referenced), the exact role and importance of different stages of joint attention has not been investigated. In this thesis, we will investigate the role of various stages of joint attention in simulated lexical grounding, using software agents that play language games in various conditions.

The structure of this thesis is as follows. In the first chapter, the concept of joint attention, its role in the evolution of language, and simulation as a method for investigating language evolution is introduced. *Language games* as a simulation environment are closer examined in chapter 2; various types of language games are reviewed and a possible adaptation to simulate various aspects and stages of joint attention is proposed. Results of simulations under various conditions (e.g., number of agents involved; number of concepts; level of ambiguity of the environment) are discussed in the third chapter. In the fourth and final chapter, these results are interpreted in terms of more general theories of language and suggestions for further research are given.

1.1 Joint attention and Theory of Mind

The term *joint attention* has been coined to describe a set of skills and interactions that emerge in infants of about nine months of age. Normally, at this age children begin to follow the gaze of their caregiver, engaging with them in social interactions etcetera. The most prominent feature in these skills and interactions is that they are *triadic*: Whereas younger children typically either have attention for a toy *or* their caregiver, the interactions of older children are usually more sophisticated and involve both the object and the other person (Baron-Cohen, 1995; Tomasello, 2000).

Carpenter, Nagell, and Tomasello (1998) categorize various forms of joint attention – like joint engagement, gaze following, and point following – into three distinct stages, namely *checking* or *sharing* attention[1], *following* attention, and *directing* attention (see figure 1). While the *follow* and *direct* stage differ in the passive versus active role of the child, the difference between merely sharing attention and following attention is somewhat more subtle. Carpenter et al. define these three stages as follows:

> Sharing attention (p.5)
> *By definition, all joint attentional skills involve infants sharing attention with a partner in some manner. We are concerned here, however, with relatively extended episodes of joint attentional engagement in which adult and infant share attention to an object of mutual interest over some measurable period of time. The prototypical example of an episode of joint attentional engagement is a situation in which adult and infant are playing with a toy and the infant looks from the toy to the adult's face and back to the toy. (…) …joint-engagement episodes are typically operationally defined by the infants' alternation of gaze between an object and the adult's face…minimally, the infant must be engaged with an object on which the adult is also focussed, then demonstrate her awareness of the adult's focus by looking to her face, and then return to engagement with the object.*
>
> Following attention (p.8)
> *It is difficult to know what infants understand of their social partners as intentional agents when they are looking to them and engaging with them in these extended periods of joint engagement. But when infants begin to follow into the attention or behaviour of others in certain specific ways, a much more compelling case can be made that they understand something about the other person as an intentional agent. In particular, infants may follow into the attention of others by following the direction of their visual gaze or manual pointing gesture to an outside object. (…) The problem is that proper interpretation (of these findings) requires some accounting of the probability that the infant will match the direction of the adult's head simply by chance.*
>
> Directing attention (p.17)
> *Human infants demonstrate their understanding of adults as intentional agents, not only by following into their attention and behaviour, but also by attempting to direct their attention and behaviour to outside entities through acts of intentional communication.*

[1] The terms *sharing* and *checking* attention are both used in the literature.

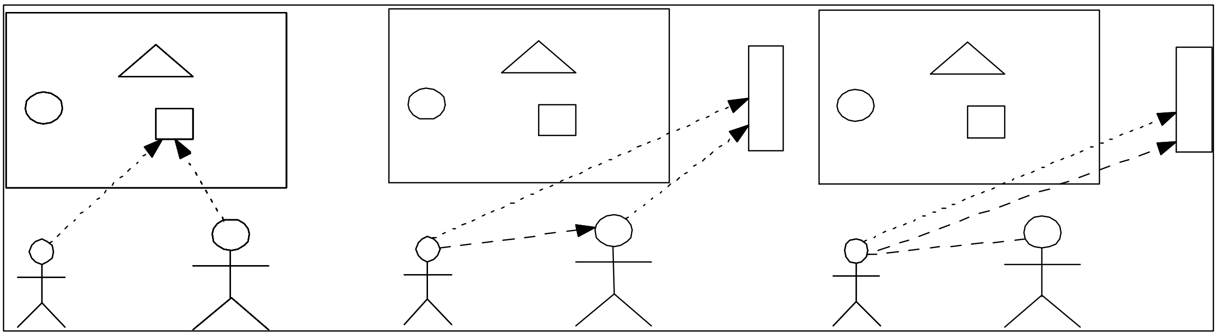


Figure 1: checking, following, and directing attention.

These definitions imply that in the *share* stage the 'third object' *is already* within the scope of the two agents (like child and adult), and in the *follow* and *direct* stage the third object is *brought into* scope, from outside. For example, while Moonwatcher and Old One were already sharing attention to the mammoth, Moonwatcher *directed* Old One's attention to the niche. Likewise, Anne *brought* Tiger into Timmy's attention. In figure 2 is shown how the scope is *extended* when the infant directs the adult's attention to the rectangle outside the whiteboard. In Chapter 2, I will discuss how these three stages can be implemented in language games.

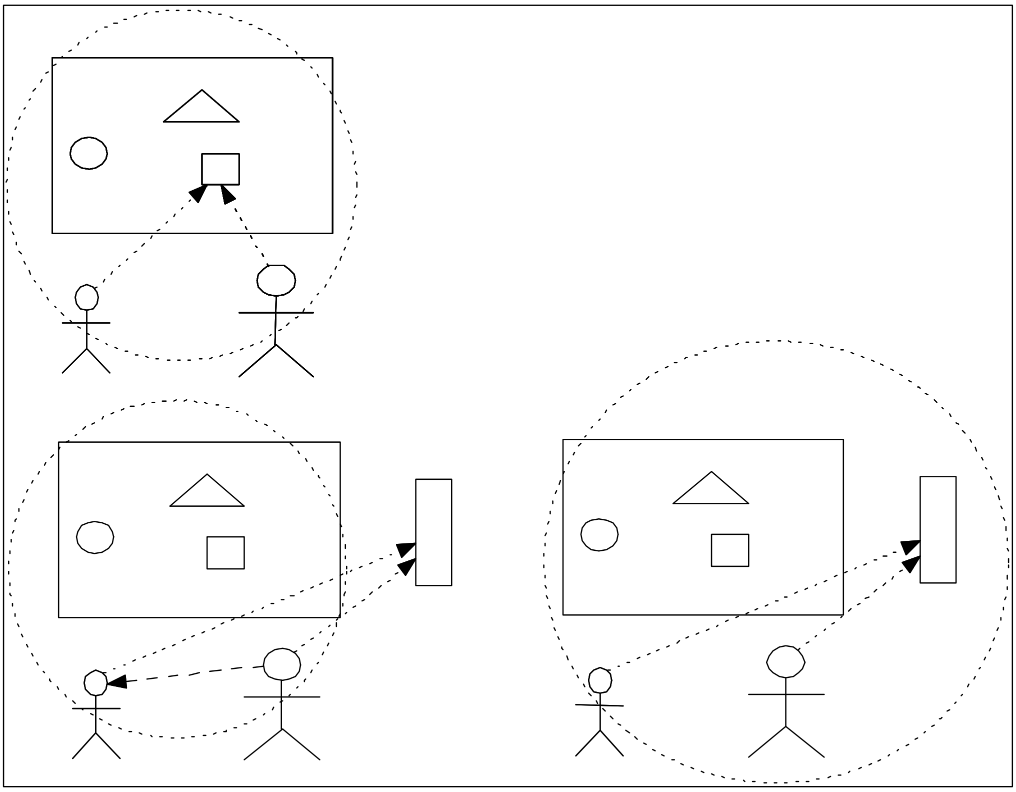


Figure 2: scope of the agents in share *versus* direct *attention.*

1.2 Joint attention, language development, and Theory of Mind

Closely related to joint attention is the concept *Theory of Mind* (hereafter ToM). Having a ToM means that one sees other actors as intentional agents like oneself, with comparable beliefs, desires, and intentions. Having a ToM is linked to *mind readers* (Sternly, 2000). Where behaviour readers like chimpanzees can interpret the behaviour (like actions and possible actions) of other actors and act according to these interpretations, mind readers have a mental model of other actors and base their expectations on this model; they *"…interpret and predict others' behaviour by the attribution of mental states, such as beliefs, desires, feelings, etc.)"* (Reboul, 2003; p.1) Only humans are believed to be mind readers and have a ToM. People suffering from the Syndrome of Asperger typically have an impaired functioning of their ToM, and often have difficulties to 'read' the mind of other people (Baron-Cohen, 1995).

It has been shown (Wimmer & Perner, 1983; Robinson & Apperlyb, 2001) that young children do not have a full-blown ToM. For example, children only pass the False Belief Test and the Opaque Context Test (see the Appendix for a description of these tests) after approximately four and five years of age, respectively. At this age, children know a considerable number of words (Bloom, 2000). Using tests like the Intentionality Detector or the Eye Direction Detector, to test various aspects of joint attention, it has been shown that infants acquire joint attention skills at approximately the same age they start to learn their first words (Baron-Cohen, 1995). They know hundreds of words at 24 months of age, long before the False Belief Test or Opaque Context Test indicate the existence of a workable ToM, as shown in Table 1, which is reprinted from Reboul (2003).

As Reboul concluded from these data, a child needs some sort of joint attention skills in order to acquire a vocabulary, but from this base ToM and language acquisition develop rather simultaneously. It is clearly not the case that a workable ToM is required before the child starts to acquire a vocabulary.

On the basis of these developmental data Reboul suggests that language evolution and evolutionary ToM development follow the same pattern. They develop in a coevolutionary way, rather than serially (ToM preceding language evolution), such that basic joint attentional skills are necessary prerequisites for both ToM development and language evolution. Malle (2002) also suggests that ToM and language have evolved “coincidentally concurrent”, as mutual escalations utilizing advances from either side, or driven by a third factor.

Table 1

Age, Language Acquisition and ToM Acquisition

Age	**Language acquisition**	**ToM acquisition**
0-9 months		ID and EDD[2]
9-18 months	Going from 6 to 40 words	SAM[2]
24 months	311 words	Development of ToM
30 months	575 words	Development of ToM
48 months	Further development of vocabulary	False Belief Test
60 months	Further development of vocabulary	Opaque Context Test

Note: data from Reboul (2003)

The hypothesis that ToM and language evolved as mutual escalations is supported by another observation in language acquisition. Although names of simple objects that play a role in the infant’s life are learned during the first years, children only use deictic relations[3] correctly at the age of three or four years, depending on the question whether the speaker’s or the listener’s perspective was taken (Pan, 2005). This also suggests a simultaneous development of language and ToM, because these simple relations do require joint attentional skills, while they are learned before passing the False Belief and Opaque Context Tests.

1.3 Joint attention and language evolution

If one compares the genetic material of humans to that of chimpanzees, one notices that they share in the order of 99 percent of their genes – the same as rats and mice, horses and zebras, or lions and tigers (Clark et at., 2003; King & Wilson, 1975). Yet the difference in cognitive skills between humans and chimps is striking, even more so considering the time in which advances were made, say from the manufacturing of stone tools to Pentium computers. There has not always been such a large difference, though. Research shows, that no hominid (like *Australopithecus*) in the first three or four million years after the separation of hominids and

[2] SAM: Shared Attention Mechanism; ID: Intentionality Detector, EDD: Eye Direction Detector. See Baron-Cohen (1995) for a discussion of these mechanisms.
[3] Deictic relations are relations whose referents depend on the speakers’ perspective, like ‘X is *behind* Y’. Children typically have difficulties specifying relations as they are experienced by another person, for example ‘to my right, and left for those of you watching at home…’.

chimpanzees had cognitive skills other than those typically belonging to the great apes. The first signs of unique cognitive skills emerged only with modern *Homo sapiens* in the last 250,000 years (Tomasello, 1999). While most probably hundreds of thousands of years separate the first *use* of fire from the first deliberately *making* of fire, the widespread use of Internet, email, and cell phones nowadays changed our way of living in more than trivial aspects in as short as ten years of time.

The puzzle is, according to Tomasello (1999), obvious: There is simply not enough time – in the biological evolutionary sense – for these cognitive skills to have emerged by genetic variation and natural selection. Tomasello does provide an answer to this puzzle. He observed that, "*...human beings are able to pool their cognitive resources in ways that other animal species are not*" (1999, p. 5). This *cultural learning* is made possible by a special form of social cognition, namely the ability to see your conspecifics as intentional agents like yourself, with similar mental lives and capacities. This requires one biological adaptation: the ability to see others as agents 'like me', that is, not just to 'read their behaviour' but also to 'read their mind'.

It should be noted that Tomasello discusses cognitive skills as a whole, and not specifically the use of language. As discussed in the previous section, Reboul (2003) and Malle (2002) suggest that Theory of Mind – seeing others as agents with a similar mental life as oneself – and the Evolution of Language use each other as stepping-stones for further development. Morales (2000) investigated the role of joint attention in language learning and found positive correlations between individual differences in joint attention development and language acquisition during the first two years. Although one should be careful to use results from language *development* as evidence for language *evolution*, this result does suggest that the ability to engage in joint attention is helpful in learning and using language, and therefore is a candidate for the biological adaptation that Tomasello had in mind.

These combined observations lead to the following model of the relationship between joint attention, language evolution and Theory of Mind development: The ability to engage in joint attention is a necessary condition for (and precedes) both the evolution of a ToM and the evolution of language. Both ToM and language develop in a co-evolutionary way, using each other as stepping-stones. Somewhere in our evolution mankind developed the ability to engage in joint attention, which made the transition from 'behaviour readers' to 'mind readers' possible, and facilitated the evolution of language.

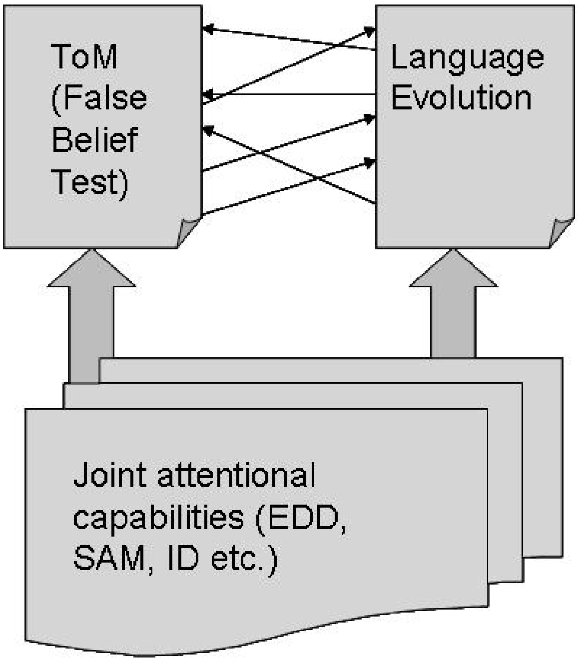


Figure 3: The relation between Joint Attention, Theory of Mind and Language Evolution

1.4 Investigating language evolution

Evolution of language can, of course, only be studied by means of indirect evidence. Especially after Darwin published *The Origin of Species* in 1859, there was a great interest in the Evolution of Language. As the textbook *Language Evolution* (Christiansen & Kirby, 2003) reported, the lack of direct evidence or other constraints on the possible explanations resulted in speculation as the primary resource for theories. This led to a ban on the subject by the Société de Linguistique de Paris, which effectively lasted until the landmark paper 'Natural Language and Natural Selection' by Pinker and Bloom (1990).

This lack of direct evidence led to the use of a wealth of research techniques from various fields, which becomes manifest when studying the chapters from Christiansen and Kirby. For example, Iain Davidson discusses how evidence for a change of diet in hominids can be interpreted as a result of the ability to plan their hunting activity (using complex cues, signs or spoken language). Natalia Komarova and Martin Nowak use mathematics to prove the logical necessity of a Universal Grammar from the observation that natural languages are considered to be more powerful than regular languages. The latter cannot be fully learned by any learning procedure; hence the human brain can only learn a restricted set of languages parameterized by a Universal Grammar.

These approaches (archaeological or mathematical) are perhaps two extreme examples of a rich variety of research methods, from psychology, cognitive science, linguistics, biology, anthropology, and philosophy. In psychology, results from language *development* research are often used as indirect evidence for language *evolution*. For example, in section 1.3 the use of deictic relations, in relation to the development of a ToM, was used as evidence for the coevolutionary evolution of language and ToM.

In this thesis, I will describe another indirect technique to investigate the evolution of language, namely that of computational methods from *artificial intelligence* and *artificial life*.

The idea of using computational methods to simulate the emergence of language has been introduced by Luc Steels. In various studies (1999a; 2001) Steels employed robots or software agents to simulate various aspects of language evolution. In the Talking Heads-experiments (Steels, 1999a) robots were used to demonstrate how they could acquire a common lexicon. The robots were located in front of a whiteboard with coloured geometrical shapes like circles and rectangles. They played *language games* that showed that agents, able of engaging in *joint attention*, established a common vocabulary of various abstract objects and properties, i.e. locations on the whiteboard like *upper-left*. In Chapter 2 I will discuss language games and the Talking Heads experiments in more detail.

1.5 Using language games as evidence for language evolution

While Steels showed that some sort of joint attention was *sufficient* for robots to establish a common lexicon of objects on a whiteboard, the model in fig. suggests that joint attentional capabilities are *crucial* in order to allow language to emerge. If a comparison between language games with or without joint attentional capabilities would reveal a strong effect (i.e., a great benefit of these capabilities), this would lend credence to the hypothesis that joint attention is a crucial predecessor – in the evolutionary sense – of language. On the other hand, if the results would be not significant or the effect was only weak, this suggests that other cognitive skills (e.g. imitation, feedback) are just as important as joint attention in this first stage, and the crucial element in language evolution may be the development of a Theory of Mind.
Of course, results should be interpreted with caution, and findings should not be over-generalised. We consider only a small aspect of language – namely the establishment of a common lexicon – in a very abstract simulation setting. Furthermore, we use these findings – from language game simulations – as an indirect evidence for language evolution. Nevertheless,

I agree with Steels (1999b) that these sort of simulations do provide valuable evidence because the emerging structures are based on the dynamics of a population of autonomous agents. To cite Steels (1999b, p.8):

> *A specific investigation consists of defining the architecture of the agents, creating a population […], and then letting the agents play a successive series of language games. In such investigations, it becomes quite natural to study language evolution. For example, one can test whether agents with a particular architecture enabling them to construct and acquire a lexicon, indeed arrive at a shared lexicon.*

Therefore I think that it is plausible that large differences in the effects of basic cognitive skills such as feedback and joint attention in these language games also suggest similar effects in evolution. That is, if joint attention would be a crucial prerequisite for the establishment of a common lexicon in language games (i.e., without joint attention, lexicon establishment is far less successful), than this indeed suggests that early hominids need to have these capabilities before more advanced language usage can emerge.

1.6 Research questions addressed in this thesis

The main research questions that will be addressed in this thesis are the following.

1. Can we find evidence in simulated language games for the hypothesis that joint attention plays an enabling role in lexicon establishment, i.e. that it facilitates a qualitative (statistically measurable) improvement of the effectiveness of the participants in these games?
2. If so, is there a statistically significant difference between the sharing, following and directing stage of joint attention?
3. Does joint attention facilitate the effectiveness of participants in an environment that is ambiguous – i.e., the interpretation of which depends on the (viewpoint of the) observer?

The *escalator*-model in fig. 3 suggests a significant effect of joint attention capabilities in lexicon establishment, at least when the environment is unambiguous. On the other hand, more complex capabilities than joint attention are needed for tasks that involve deictic observations. This suggests less effect of these capabilities in environments that are ambiguous, i.e., where certain objects are observed differently, depending on the (viewpoint of the) observer. To resolve these ambiguities, it is necessary to ‘stand in someone else’s shoes’, i.e., have a Theory of Mind. Therefore, the model suggests a lower effect, or none at all, in ambiguous situations. It might even be the case that in ambiguous environments, joint attention hinders – rather than facilitates – language establishment.

For example, in the observational game there are typically multiple concepts that may be the topic of the current language game, such as the colour of the object under consideration, its size or its shape. The weight values of all of these concepts will change as a result of the language game, whereas in joint attentional games typically fewer – or even only one – concept will be selected, such that a mismatch will have a relatively greater impact on the weight values. Of course, a match also has a greater inpact: It depends on the parameter settings in the experiment, e.g. the learning and failure rates, whether positive or negative effects will dominate

1.7 Sub questions

Apart from these questions regarding language evolution, these simulations can provide answers to some interesting sub-questions regarding language acquisition, if we use an iterated learning game setting to simulate an adult-child setting where the child needs to learn the adults’ lexicon[4]. In this setting, we can look for support for observations that joint attention plays a crucial role in language *acquisition* (i.e., the situation where a child learns a language from a caregiver, rather than two people try to establish a common grounding of objects and properties) as found by Morales (2000). Furthermore, we can look for support for the observation that *following* the child’s attention leads to faster language learning than *directing* the child’s attention (Tomasello & Farrar, 1986), and even simulate hypothetical situations, such as language grounding in a situation where one of the players in the game is able to direct one’s attention, but lacks the ability to follow attention, e.g. due to brain damage.

[4] It should be noted that results from these sub-questions, however interesting, are not the main focus of this thesis research. The language games are designed to test the role of joint attention in establishing a common lexicon, rather than language acquisition, and while interesting observations will be discussed in this thesis, they should be interpreted with caution.

2 Model, simulation, and empirical research

2.1 Language Games

As discussed in section 1.2, an often-used method to investigate the evolution of language is computational modelling of language development. Typically, the emergence of a common language is modelled with a population of agents that play so-called *language games* (Steels, 1996). In these games, a population of agents tries to develop a grounded lexicon using communicative actions in a particular environment (e.g., a whiteboard with coloured geometrical figures). Such a language game is typically played between two agents; one of them trying to label the object or feature of objects the other is looking at.

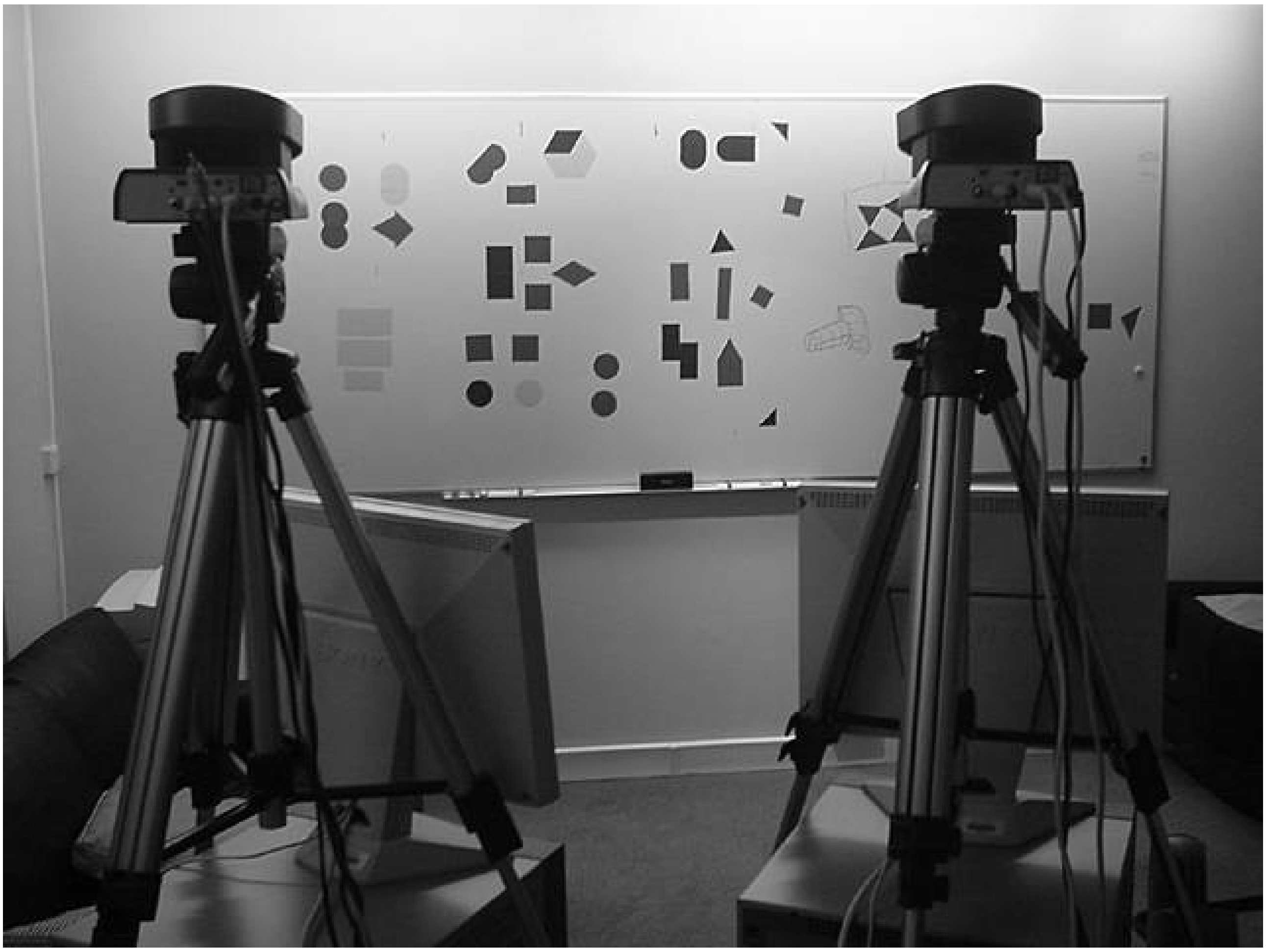


Figure 4: Two robots playing a language game with geometrical objects on a whiteboard (picture from Talking Heads website, talking-heads.csl.sony.fr).

In the Talking Heads experiments, as introduced in section 1.5, the agents were robots with cameras, capable of zooming in on the whiteboard and detecting the other robot's zooming, mimicking ‘looking’ or ‘pointing’ at a certain region of the whiteboard (figure 4). In the *guessing games* (the language games played in the Talking Heads experiment), the first robot (the *talker*, A) would select one of the objects on the whiteboard, search its memory for the label with the highest weight value for that object, and communicates that label or utterance.

The second robot (the *hearer*, B) now searches its memory for the object that matches the utterance best and ‘points’ to that object. Robot A now interprets this pointing and gives feedback on the outcome. Based on this feedback, the weights of the labels are changed, and the robots engage in other games with possibly other robots. After a certain amount of games, a certain label becomes dominant in the population of agents.

Vogt and Coumans (Vogt & Coumans, 2003) describe three types of language games, using joint attention (observational game), corrective feedback (guessing game) or no feedback or joint attention at all (selfish game). Observational games were described in Oliphant (1997); guessing games in Steels and Kaplan (2002), whereas Smith (2001) introduced selfish games. In this thesis, these games were organised at follows (see also the pseudo code of these language games in the Appendix).

In all types of games, the population consists of two agents, taking alternation terms as a *talker* and a *hearer*. The agents try to develop a shared lexicon for *attributes* of objects, e.g. their colour or size. Both agents ‘see’ four objects that have two such attributes each. In the *guessing game*, the talker selects an attribute like ‘large’ or ‘red’ out of a fixed set, it searches its database for the label with the highest weight for this concept and communicates this label. The hearer tries to associate this label with an attribute and selects an object out of the four available that has this attribute. The hearer communicates this object and the talker gives corrective feedback, indicating that the object has or hasn’t the attribute that was referred to. Both agents adapt the weights of their labels based on this outcome.

In the *observational game*, the talker again selects an attribute, selects an object that has this attribute, and communicates the object to the hearer. The hearer tries to associate the attributes of this object with labels and communicates the label with the highest weight value to the talker. Both talker and hearer adapt the weight values according to this communication. The *selfish game* is modelled as a special case of the observational game, where no specific object is communicated.

The language games can be visualised by the following metaphor of a mother (M) and her child (C), both sitting on a blanket with a number of toys, say a red ball, a green ball, a red doll and a green car. In the guessing game, M names a colour ('red') and C points to the red ball, after which M gives *feedback* on the object that C had chosen. In the observational game, M picks an object and presents it to C (thus *sharing attention* of this object), and C must name an attribute of that object (for example colour). This also happens in the selfish game, but no object is presented explicitly. These games are then repeated in a different environment, i.e., with different toys.

Using the observational game, two extensions were programmed by us to model joint attentional skills. The hearer can either ask for another example of an object that has the attribute under consideration, or select such an object for consideration by the talker itself. These two extensions are a model of the hearer's capability to follow, respectively direct the attention of the talker. Using these extensions, we can arrange the language games by the type of joint attentional skills they model. Note that the Guessing Game is based on feedback, rather than joint attention.

Table 2.

Languages Games and their Correspondence to Joint Attention Stages

Game	**Joint Attention**	**Level**	**Acronym**
Guessing Game	None, feedback	N/A	GG
Selfish Game	None at all	0	SG
Observational Game	Check attention	1	OG
OG with follow attention	Follow attention	2	JA1
OG with direct attention	Direct attention	3	JA2
OG with follow and direct attention	Follow and direct	4	JA3

While I think that the ability to employ *check attention* is a *sine qua non* for both the ability to follow and direct attention, I intentionally distinguish between the situation where the agent acquires direct attentional capacities next to the capability to follow one's attention (level 4) and the situation where this follow capacity is lacking (level 3), which could model some hypothetical brain damage.

2.2 Language games versus language learning

As discussed in De Vogt and Coumans (2003), children often learn words by *associative learning*, where both words and their referents are presented simultaneously, requiring some sort of joint attention. Nevertheless, in some Eastern cultures parents do not speak directly to their children before they already know some words. The children must grasp the meaning of the parent's utterances without explicit cues to their referents (Lieven, 1994). An alternative for associative learning is *reinforcement learning*, where children receive feedback on their language use. This learning strategy is sometimes observed in Western middle class families (Lieven 1994).

These different strategies can be related to, respectively, the observational, selfish, and guessing game. Typically, children who (are forced to) employ strategies that do not involve joint attention or feedback learn their first words slower than those who do use these strategies (Lieven 1994).

2.3 Reported experimental results

De Vogt and Coumans (2003) discussed experimental results for the selfish, guessing, and observational games in terms of the communicative success. They averaged over 10 simulations (with different random seeds), each consisting of 50,000 language games, with a population size of 10 agents taking alternative turns. The *communicative success* of these simulations was defined as the number of successful games averaged over the last 100 games; the *coherence* was defined as the average rate in which each agent would produce the same label to express a meaning. Furthermore, entropy measures as *specificity* and *consistency* were measured in these simulations.

The results show that all types of games would eventually reach a 100% communicative success rate, where both guessing and observational games would reach this level at 10,000 games and the selfish game would converge much slower. The guessing and observational games would reach a coherence rate of 100%, where the selfish game would stick at a coherence rate of 5%. In their discussion, De Vogt and Coumans suggested that the more rapid convergence in both guessing and observational games has a relation to the *fast mapping phenomenon* described by Carey (1978), where children learn many novel words within only one or two exposures.

Furthermore, they stressed that the way joint attention and corrective feedback is modelled is not very realistic.

2.4 Implementation of joint attention in the observational game

In this thesis, the three stages of joint attention (check attention, follow attention and direct attention) were implemented in the observational game as follows. Depending on the stage of the hearer in the game, the hearer had a sort of *toolbox* with methods to narrow the set of possible attributes the talker was referring to. For example, when the talker communicated a red triangle, the set of possible attributes for the hearer was {red, triangle-shaped}. The toolbox consisted of two additional methods:

1) to ask for another object that also has the attribute the hearer wants to communicate about
2) to search for, and communicate another object and ask whether this object also has the particular attribute

The first method requires follow attention (or level 2 according to table 2.1), the second method requires direct attention (or level 3). A level 4-agent has access to both methods.

For example, if the set of applicable attributes was {red, triangle-shaped}, a level 2-agent could ask for another object, and receive a red circle, in which case the *intersection* of {red, triangle-shaped} and {red, circle-shaped} would result in ‘red’ as the only possible attribute left. A level 3-agent could communicate a blue triangle and would receive a negative answer. In this case the *disjoint* of {red, triangle-shaped} and {blue, triangle-shaped} would result in ‘red’ as the only possible attribute.

Note that in this implementation I am using an abstract notion of follow and direct attention. Instead of the actual pointing and gaze following as described in Carpenter et al. (1998), I use a particular feature of follow and direct attention, namely that it refers to an object which was previously *not in scope*, in contrast to the object that was subject of the language game. The new object was *brought* into scope, either by the talker (level 2) or the hearer (level 3). Using the Mother and Child metaphor, in the level 1-game the object was assumed to *be* a part of the shared attention (for example because M took it from the blanket and gave it to C), in the level 2 and 3-games the new object was *brought* into attention because either M or C pointed to that object.

2.5 Simulation settings

The agents and the language games were implemented using the agent programming language and interpreter 3APL (to be found at http://www.cs.uu.nl/3APL). This language allows the user to specify agents in terms of their beliefs, desires and goals, and use practical reasoning rules and Prolog clauses to reason on their beliefs. The agent interpreter explicitly deliberates on the agent goals, reasoning rules etcetera.

Although the use of an agent programming language (rather than JAVA or C++) facilitates the specification and coding of the participants in the language game as goal-driven agents, the actual implementation of the 3APL interpreter – which is still in an experimental stage – limits the number of games that can be played due to memory constraints.

2.5.1 Method

Series of simulations were run with guessing games, selfish games, and four types of observational games, as shown in table 2. Each simulation consisted of 100 language games and was repeated ten times with different random seeds. In each language game the environment consisted of four objects.

In the first series, there were two agents, who alternatively took the role of the *talker* and the *hearer* in the language games. Each of the four objects in the environment had two distinct properties; this was the baseline condition. In the second series, each object had three properties instead of two; in the third series there were three agents instead of two, who played language games in a tournament-like setting (i.e., every agent played against both other agents in either role). In the forth series, the level of ambiguity in the environment was varied with respect to the baseline condition. In the fifth series the lexicon was fixed for one agent (the adult) and in the language games the other agent (the child) was always the hearer.

Parameters and settings were comparable to those in the simulations run by De Vogt and Coumans (2003), except that the number of language games and participating agents was limited to 100 language games and, at maximum, three agents, due to the constraints of the 3APL interpreter. In the simulations of De Vogt and Coumans, the number of language games was limited to 50,000 and the number of agents varied from two to twenty. In section 3.2 I will illustrate how a typical simulation would look like if the number of games were increased. In Appendix 3 a summary of the relevant parameters in both experimental settings is presented.

Note that in these language games the naming of objects is arbitrary, i.e., there is no preferred mapping between object and label. It is questionable whether this is always the case in language development. For example, Köhler (1929) designed an experiment in which participants had to map two labels (‘booba’ and ‘kiki’) to either a rounded or an angular shape. More than 95 per cent of the participants chose ‘booba’ for the rounded shape and ‘kiki’ for the angular shape.

2.5.2 Coherence rate

De Vogt and Coumans (2003) used four measures to indicate the result of the series, namely coherence, communicative success, specificity and consistency defined as follows:

Communicative Success: the number of correctly played games averaged over the past 100 games or less when no 100 games have been played yet.
Coherence: average rate in which each agent would produce the same word to express a meaning.
Specificity and **Consistency**: entropy measures specifying the amount of polysemy respectively synonymy in the lexicon.

In this thesis, I use a simplified measurement, namely the coherence rate of the language games. This coherence rate is calculated after a certain number of games (or after the total simulation run) to indicate the current coherence of the agent’s lexicons. It is calculated as follows:

For each attribute, the weights of all possible labels for agent A and B are multiplied. Then, the largest absolute difference of the thus obtained values is called the *coherence measure* for this attribute. If three agents are playing language games, the procedure is followed for the weights of A versus B, A versus C and B versus C and these results are averaged.

As an example, take the following weights after 100 language games in table 3. For attribute e1, there are two labels for which the agents have non-zero scores, namely *vnkt* and *vwbg*. The pair wise product of the weights for each label for A, B and C are multiplied, and their average is taken. The largest absolute difference for any of the labels is 0.1254 (for *vwbg*) - 0.010 = 0.1244.

Table 3

Example of weight distribution

Attribute	Label	Agent A	Agent B	Agent C	A vs. B	A vs. C	B vs. C	Averages	Coherence
e1	vnkt	.03	.01	.07	.0003	.0021	.0007	0.0010	.1244
	ywbg	.42	.23	.43	.0966	.1806	.0989	0.1254	
e2	pkcj	.47	.56	.86	.2632	.4042	.4816	0.3830	.3830
e3	pkcj	.82	.30	.10	.2460	.0820	.0300	0.1193	.0308
	koff	.01	.56	.78	.0056	.0078	.4368	0.1501	

2.5.3 Coherence development over time

Figure 5 shows an example of coherence development in a typical simulation run of 100 language games, with six abstract concepts e1 to e6, grouped in sets of three per attribute. Initially, all possible label-concept associations start with zero weight. After some successes, the coherence for certain concepts increases, but there is still much competition between various labels and concepts. Only after 40 games or so, the coherence steadily increases, except for e3 and e5. These concepts share the same label (i.e., they are homonyms) and a frequent occurrence of e3 in the environment lowers the coherence rate of e5, and vice versa. This effect can be compared to e.g. the use of the term ‘chair’ by someone who has just been elected to a representative body.

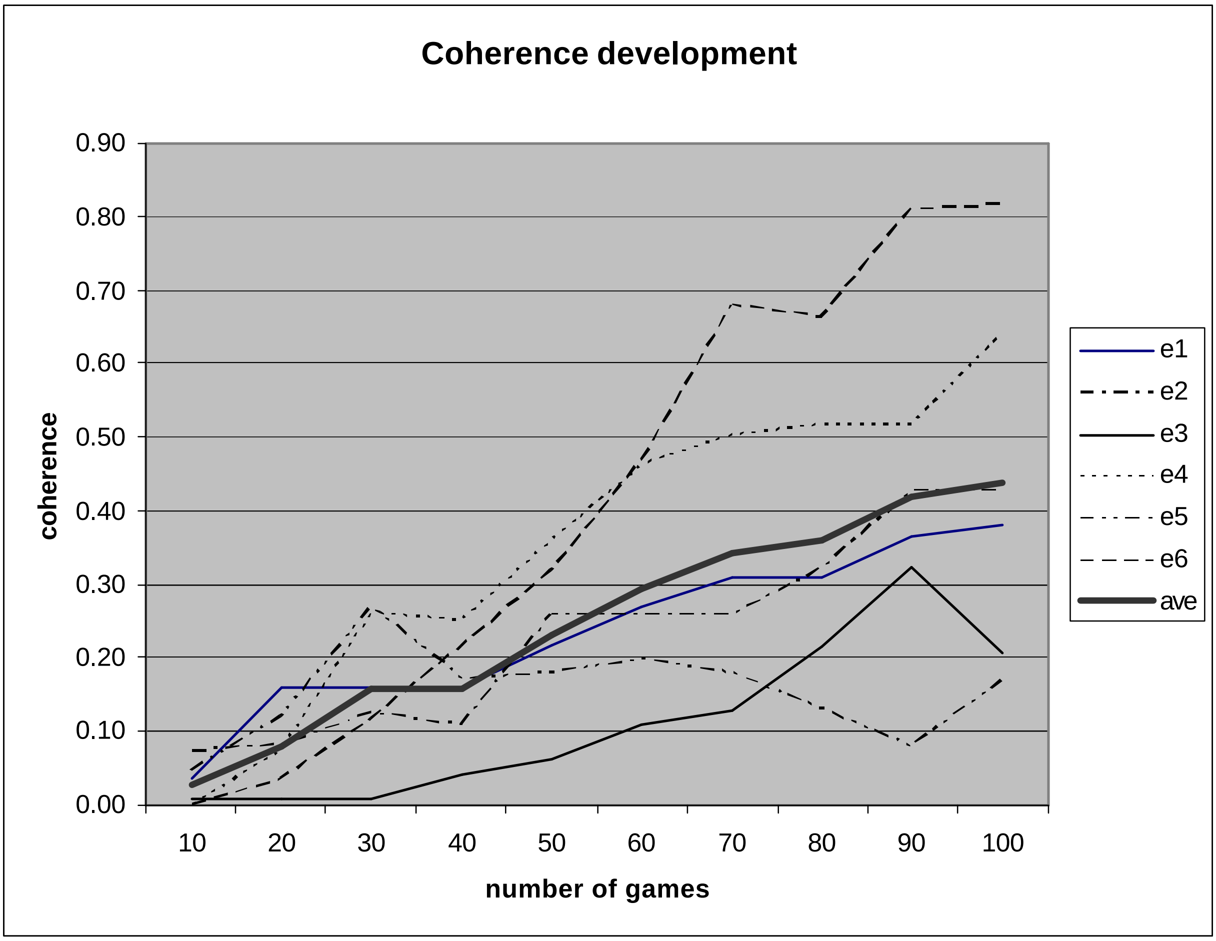


Figure 5: Example of coherence development during 100 language games

Table 4 shows how the coherence rate in this situation would have developed if the number of language games would be doubled. While the coherence of most labels has become greater, the use of e3 and e5 as homonyms is rather stable, although their relative weights may change over time.

Table 4

Coherence development after 100 and 200 games

	e1	e2	e3	e4	e5	e6	Mean
100 games	.3795	.4260	.2064	.6384	.1680	.8160	.4391
200 games	.7040	.5325	.3584	.6570	.0700	.6120	.4890

Apart from these homonyms, it is also possible for synonyms to emerge in these language games, comparable to e.g. ‘chair’ and ‘seat’.

3 Results

3.1 Preliminaries

In the experiments, I used language games with two agents, and objects with two properties (each with three possible values) as a baseline condition, and varied the number of agents, number of properties and level of ambiguity in the environment. For each type of language game, 10 simulation runs were played with different random initialisers. After a complete run of 100 language games, the coherence rates were calculated per attribute and averaged. These coherence rates of the 10 simulation runs were analysed using a ANOVA with game type as the between subject factor and coherence rate as the dependent variable. After that, similar simulation runs were played with 3 participating agents (rather than 2), 3 distinct attributes of each object (instead of 2), and with an environment which was slightly, resp. moderately ambiguous. The results of these simulation runs were then compared with the baseline simulations using the following ANOVAs:

- a 6x2 ANOVA with game type (GG, SG, OG, JA1, JA2, JA3) and number of agents (2/3) as between subject factors, and coherence rate as dependent variable;
- a 6x2 ANOVA with game type and number of attributes (2/3) as between subject factors, and coherence rate as dependent variable;
- a 6x3 ANOVA with game type and level of ambiguity (none/slight/moderate) as between subject factors, and coherence rate as dependent variable;

Furthermore, simulation runs were played which used fixed roles in an iterated language game, where the labels were fixed for the talker. The results were analysed and the coherence rates compared to the coherence rates in the baseline simulations using a 6x2 ANOVA with game type and role (variable/fixed) as between subject factors, and coherence rate as dependent variable;

In the simulations I used abstract attributes and abstract objects in the environment: The attributes were denoted as A1, A2 and A3 and the values these attributes could take were denoted as e1 to e9, grouped by three (i.e., A1 could take values e1 to e3 and so on).

In the remainder of this chapter, I will first describe results of the baseline setting (2 agents, 2 distinct attributes, unambiguous environment) for all types of games. In section 3.3 I will describe what happens when three attributes are used, instead of two. In section 3.4 I will describe simulations with three agents, and in section 3.5 simulations in ambiguous environments are discussed. In section 3.6 I discuss iterated learning games.

3.2 Language game results

In this section, language games are played with two agents in an environment with four objects, each having two different attributes. The average coherence rates are as shown in table 5 and figure 6. The results are significantly different ($F(5,54 = 20.64, p < .001)$. The guessing game scores higher than the selfish and standard observational game, comparable to the follow attention and direct attention enhancements, and lower than both enhancements combined.

In this rather simple setting, the joint attentional enhancements have a great impact on the coherence rate. The scores are doubled, compared to the standard observational game, and almost six times as large as in the selfish game. Nevertheless, feedback – as used in the guessing game – is still a successful strategy in this setting.

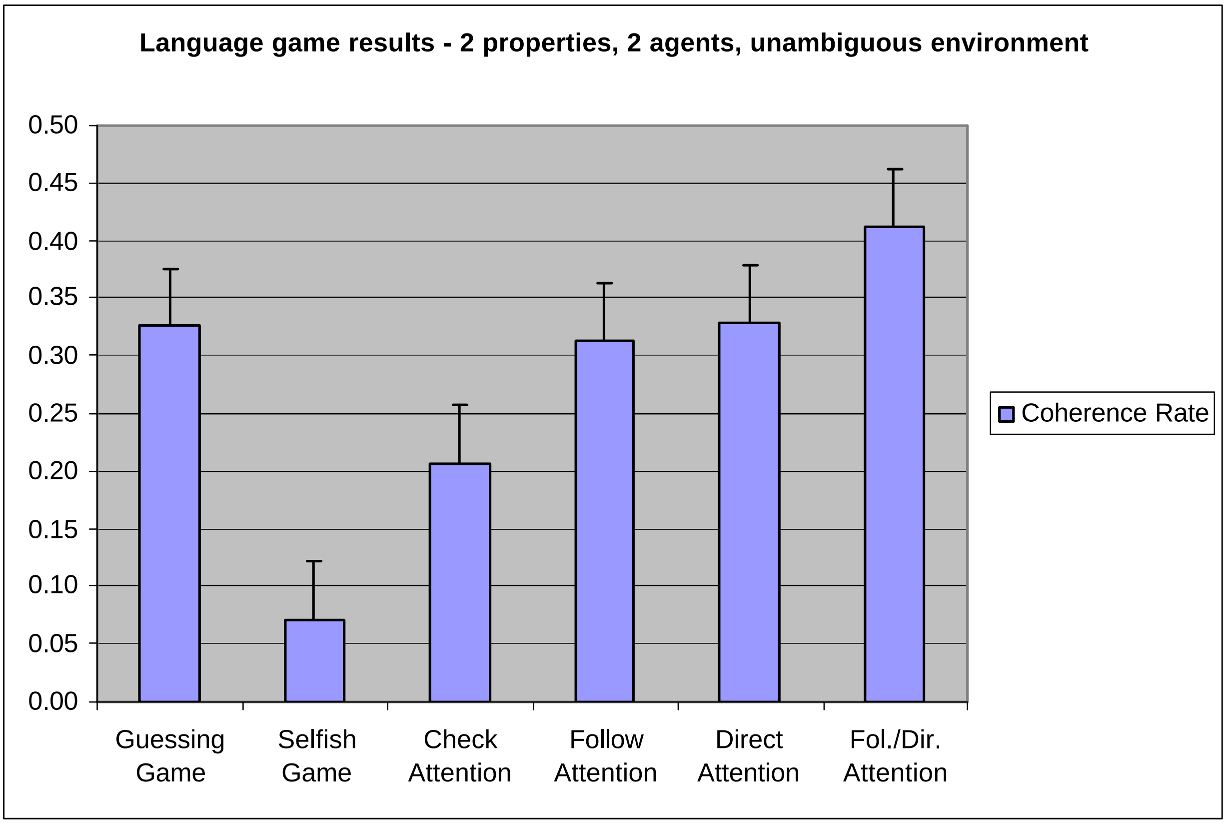


Figure 6: Scores in the baseline condition

Table 5

Means and standard deviation in baseline condition

	GG	SG	OG	JA1	JA2	JA3
Mean	.3257	.0712	.2067	.3122	.3288	.4118
St.dev	.1477	.0317	.0777	.0475	.0653	.0786

After presenting these results, I will summarize the results for each game type in the following sections. In the appendix, all results are given.

3.2.1 Guessing Game

In the guessing game, the primary mechanism used for lexicon development is feedback. The coherence rates in these games vary a lot between the simulation runs: the data shows a minimum of .0844 and a maximum of .5764. This variation is also visible in the relatively high standard deviation of .1477. If we take a closer look at the results, we can see that there are typically some labels that have a high coherence rate, and some that have a low coherence rate or even none at all. This is caused by the *guessing* nature of this game: there is no correction on the communicated labels, only an acknowledgement of success or failure. There can be a lot of mismatches between the agents before they agree on a common label. Therefore, the weights for some games are still low after 100 language games.

3.2.2 Selfish Game

In the selfish game, there is neither feedback nor joint attention, the only mechanism used is associative learning. As a consequence, the agents need a large amount of games to associate labels with concepts, purely on a statistical basis. There is large confusion regarding the labels of the attributes, and little coherence: most scores are low. Only a few labels dominate and are used for most of the concepts. The overall results are probably just a little higher than chance levels after 100 games, which matches the results of Vogt and Coumans (2003) how found considerately lower coherence rates for selfish games than for guessing or observational games.

3.2.3 Observational Game

In the observational game there is no feedback, but there is shared attention; both agents do know which object is the topic of the game. It is, however, not possible to resolve which attribute of that particular object is meant by the talker. This is reflected by the results: there are different labels used to distinguish between the different values of one attribute but only one attribute is favoured. Throughout the simulation runs there is a preference for attribute A1 (i.e., this attribute gets higher scores) due to the Prolog implementation: when multiple goals are satisfiable, e.g. when one item out of a set of two is to be selected, the Prolog implementation always selects the last one. This effect also causes the high standard deviation in this game.

3.2.4 Joint Attention enhancements

The enhancements to the observational game, adding the possibility to further specify the attribute under consideration, have a considerable influence on the coherence rate. While in the observational game, where only *share attention* is used, there is typically one label for a value of A1 and A2 that is used simultaneously, the labels are more specific in the games with follow, direct, and most notably the combination of follow and direct attention. The results of the *direct* and *follow attention* conditions are comparable to each other. In the combination of follow and direct attention, the joint attention enhancements caused almost a doubling of the average coherence rate, with respect to the standard observational game.

3.3 Using three attributes

When the objects become more complex and have three attributes instead of two (with, again, three possible values for each attribute), the coherence results drop more than would be expected. The number of attributes increases with 50%, but the coherence rate is at best halved and drops to slightly more than a third of its original value in the observational game. Nevertheless, the use of additional joint attentional methods to specify the attribute under consideration does mitigate the decrease compared to the observational game. There is a significant interaction between game type and number of attributes ($F(5,108) = 4.64$, $p < .001$).

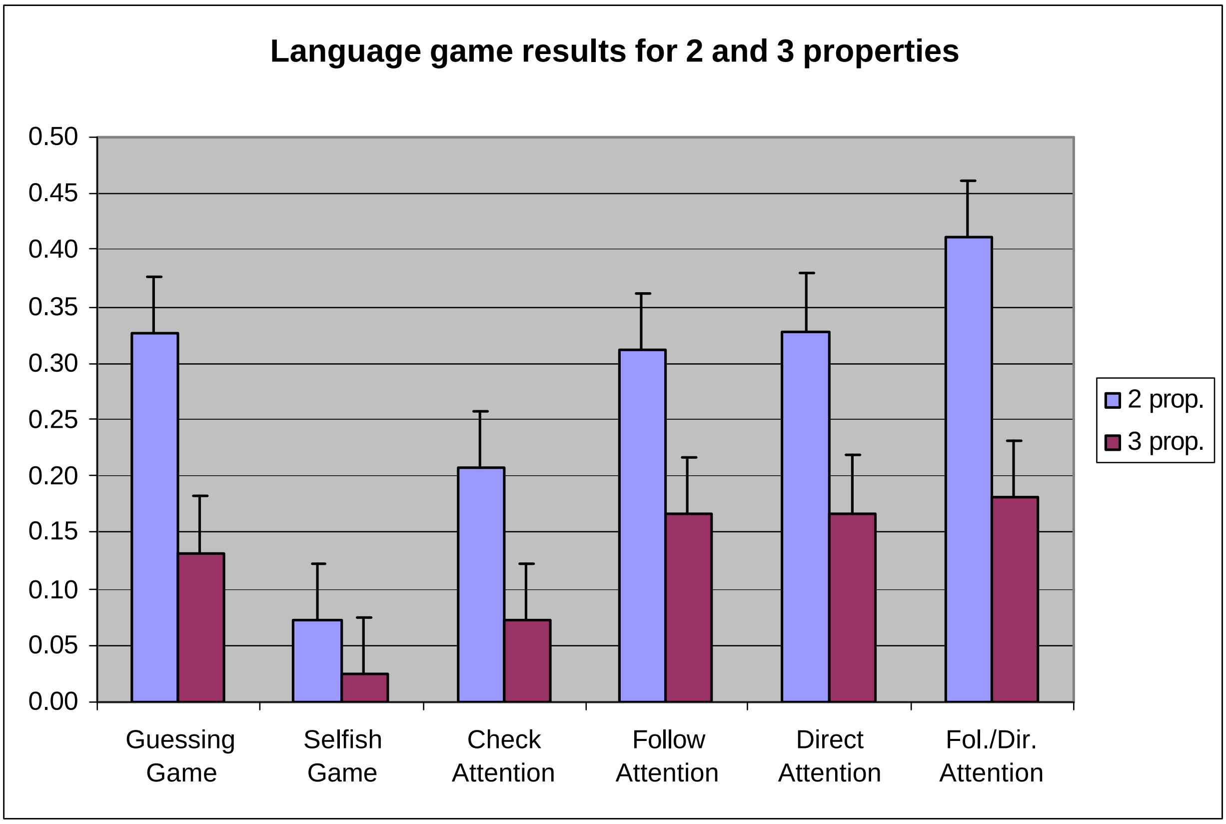


Figure 7: Scores with 2 and 3 properties

Table 6

Means and standard deviation with three attributes

	Attributes	GG	SG	OG	JA1	JA2	JA3
Mean	Two	.3257	.0719	.2067	.3122	.3288	.4118
	Three	.1321	.0251	.0729	.1658	.1674	.1809
St.dev	Two	.1477	.0317	.0777	.0475	.0653	.0786
	Three	.0593	.0107	.0344	.0276	.0422	.0381

3.4 Using three agents

If the language games are enhanced to three participating agents, taking alternative roles in a tournament-like setting, we see a decline in the guessing game but no dramatic changes in the selfish, observational, and joint attentional games. The interaction between game type and number of agents was not significant ($F(5,108) = 1.15$, $p > .1$).

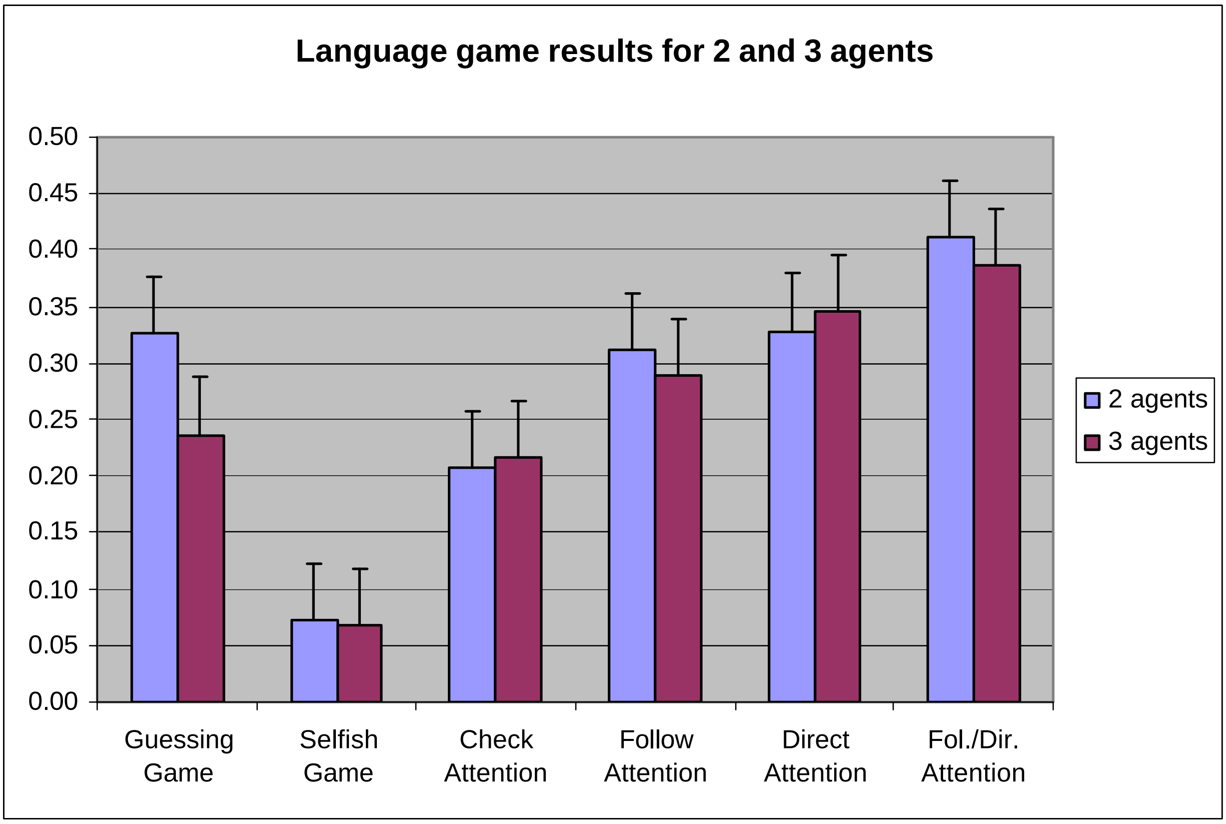


Figure 8: Scores with 2 and 3 agents

Table 7

Means and standard deviation with three agents

	Agents	GG	SG	OG	JA1	JA2	JA3
Mean	Two	.3257	.0719	.2067	.3122	.3288	.4118
	Three	.2365	.0674	.2162	.2882	.3459	.3861
St.dev	Two	.1477	.0317	.0777	.0475	.0653	.0786
	Three	.1179	.0307	.0674	.0714	.0758	.0695

3.5 Using ambiguous environments

We also varied the level of ambiguity in the environment, determining the probability that the agents had different observations of an object. In the guessing game and in the observational game without enhancements, there is not much difference in the coherence rates. In the selfish game the scores drop even with a slightly ambiguous environment, but most striking is the collapse of the scores of the joint attentional enhancements. In a moderately ambiguous environment, these games scored about the same as the non-enhanced observational game: the enhancements are of little use in an ambiguous environment. The interaction between game type and ambiguity level is significant ($F(10,162) = 2.31$, $p < .05$).

Table 8

Means and standard deviation with ambiguous environments

	Ambiguity	GG	SG	OG	JA1	JA2	JA3
Mean	None	.3257	.0719	.2067	.3122	.3288	.4118
	Slightly	.3314	.0676	.2052	.3103	.3206	.3720
	Moderate	.3425	.0693	.2046	.2175	.2522	.2253
St.dev	None	.1477	.0317	.0777	.0475	.0653	.0786
	Slightly	.1303	.0339	.0779	.0415	.0580	.0746
	Moderate	.1777	.0315	.0882	.0740	.0666	.0672

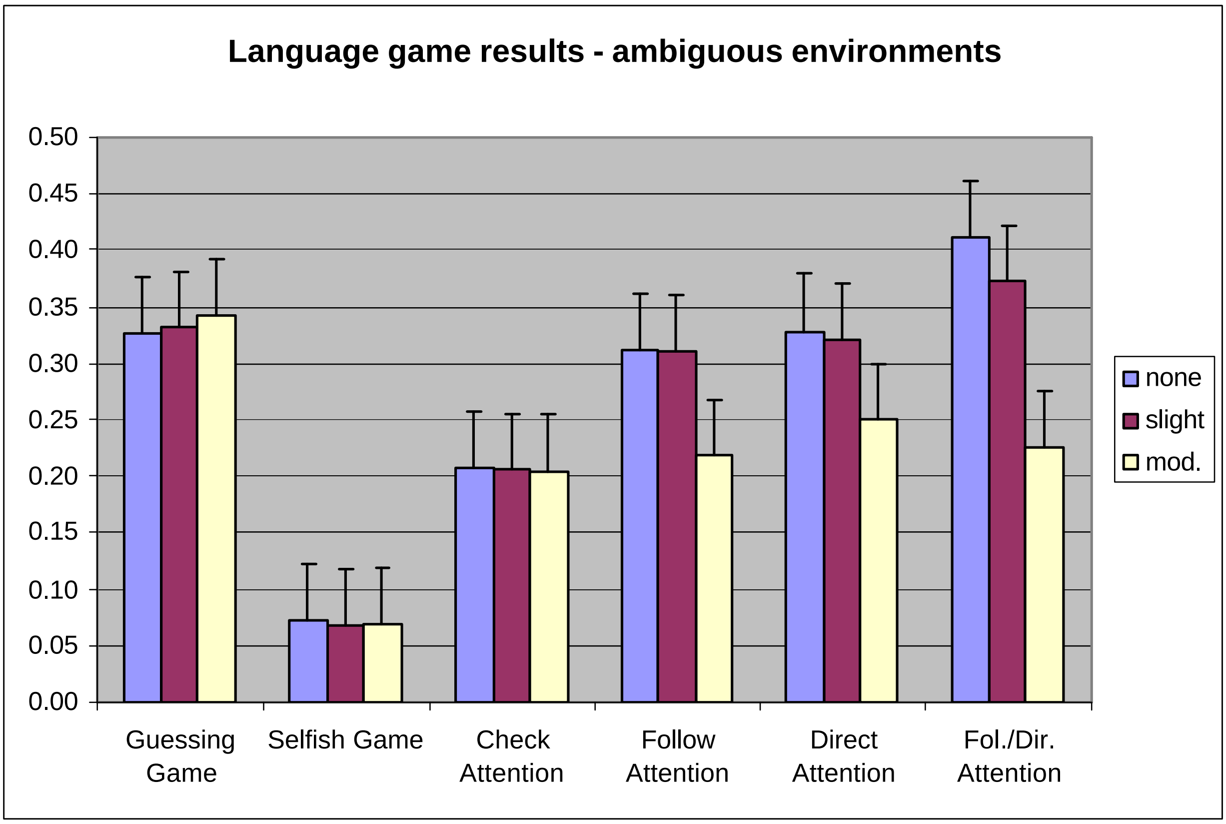


Figure 8: Scores in environments with no, slight, and moderate ambiguity

3.6 Iterated learning game with fixed labels and fixed roles

In the iterated learning game, one agent is always the talker and the other the hearer. Actually, the talker plays a ‘teaching’ role with the hearer as pupil. For the talker, the labels are fixed; the hearer has to learn these labels. There is no coherence rate because the labels are fixed for the talker; the scores are calculated using the weights of the matching labels for the hearer.

The scores for the guessing game do not rise with respect to the grounding games, in contrast to the other game types. There is a significant interaction between game type and language task ($F(5,108) = 105.00$, $p < .001$).

Table 9

Means and standard deviation with variable and fixed roles and lexicons

	Roles	GG	SG	OG	JA1	JA2	JA3
Mean	Variable	.3257	.0719	.2067	.3122	.3288	.4118
	Fixed	.2992	.2048	.4515	.5707	.6120	.6498
St.dev	Variable	.1477	.0317	.0777	.0475	.0653	.0786
	Fixed	.1412	.0243	.0328	.0311	.0688	.0409

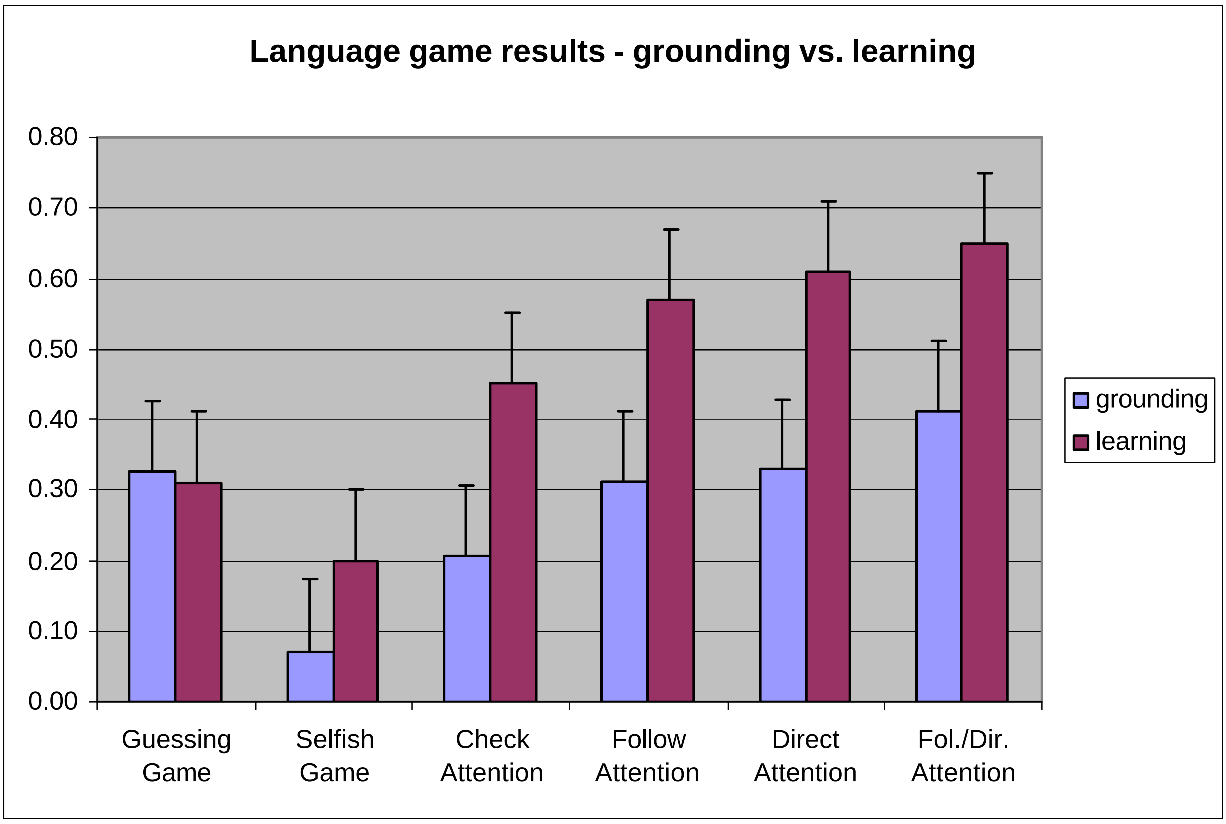


Figure 9: Scores in grounding and learning games

3.7 Discussion of these results

From these results, it is clear that the use of these joint attentional enhancements has a large impact on the observational game in unambiguous environments. While feedback can be an alternative in a very basic setting, it is a far less successful strategy when the number of attributes is increased. While the advantage of the joint attentional enhancements completely disappeared in the moderate ambiguous environments, the results for the guessing game were remarkably stable.

This effect might be attributed to various causes. One partial explanation might be that mismatches due to the ambiguity are compensated by the prevention of false acknowledgements in other cases. For example, suppose that the environment is a red circle, red triangle, and white circle in the unambiguous condition. If the first agent's label for 'red' matches the second agent's label for 'triangle', there can be a false acknowledgement if the second agent looks at the red triangle. If, on the other hand, the first agent would see a *white* triangle instead of a red one, this ambiguity would prevent such false acknowledgements.

The same explanation may hold for the selfish and observational games. While it is difficult to discriminate between the attributes in these games (e.g., a label can be associated with either the colour or the shape of the object), the weights of both associations will change. In ambiguous environments, the effects of a mismatch might be balanced. This is typically not the case in the conditions with joint attentional enhancements where it is easier to denote the concept that is the topic of the game.

The guessing game scores in the iterated learning game are much lower than the scores in the observational game with or without joint attention enhancements. This supports the observations of Morales (2000) that joint attention plays a crucial role in language acquisition, as discussed in chapter 1. On the other hand, the *follow* and *direct* attention enhancements did not differ significantly, thus there is no support for the observation by Tomasello and Farrar (1986) that *following* the child's attention leads to faster language learning than *directing* the child's attention.

4 Discussion

4.1 General discussion

What is it that makes us human? Amongst various possible answers to this question, the ability to use language stands out as one of the most important features that separate mankind from other hominids. The use of language is universal in all societies and tribes of mankind. It is certain that man acquired the ability to use language somewhere during our evolution. Why, when, and how this happened, is subject of continuous research in branches as diverse as psycholinguistics, philosophy, evolutionary anthropology, and artificial intelligence. This diversity in research approaches is necessary in order to collect evidence from various sources, because no direct evidence (e.g., written texts or speech recordings) is available on this subject.

While consensus is reached upon many questions in this diverse research community (e.g., that the subject must be approached from various disciplines) some open questions remain in the field, as discussed by Christiansen and Kirby (2003, p.14). For example, it is an issue whether grammatical structure and its constraints emerged as a consequence of an evolved innate grammar as a cognitive structure that enables humans to use language (a *Universal Grammar*), or whether it emerged through cultural transmission.

The main argument for this genetically determined Universal Grammar, rather than more general cognitive skills, is the *paradox of language acquisition* (Jackendoff, 1997). The paradox is as follows. Children are able to form grammatically correct novel sentences, without any formal training of these grammatical rules. Although they suffer from the so-called *poverty of stimulus*, i.e., they do not encounter all possible correct combinations of words that would enable to learn these rules in a sort of associative way (and furthermore, they hear *incorrect* combinations without the explicit notion that they are incorrect), they can correctly use these rules without explicitly knowing them.

Chomsky posed a solution to this paradox by introducing the concept of a Universal Grammar (1968), loosely described as follows. There are universal grammatical rules, available to the child as a sort of instinct, and when learning a language the child only has to ‘fine-tune’ these rules to accommodate the language under consideration. An example is, that children do not need to learn *that* sentences are formed using the abstract concepts ‘noun phrase’ and ‘verb phrase’, because these concepts are part of the Universal Grammar. A child ‘knows’ by instinct that a sentence has an agent and an activity. Only *how* a sentence is built in a particular language (e.g. which word order) is to be learned by the child.

However, as MacWhinney (1998) pointed out that there is no evidence whatsoever of a discrete moment in the child's language development in which the child 'sets some crucial parameter'. Davis (1947) shows that children can learn language after they have been isolated for as much as 6 years. Tomasello refuted the existence of a Universal Grammar, and posed that language universals result from human cognitive and social universals rather than an innate framework, and that we should look at other human skills to explain language (1995). Nevertheless, the question whether language developed as a separate module or as a result (or by-product) of other cognitive developments, is still subject of hot debates (see for example the articles in Christiansen & Kirby, 2003).

Another issue is whether spoken language gradually evolved out of primitive gestures of hominids, or whether its origins lie exclusively in human evolution (i.e., spoken language is *qualitative* different from signs and calls) is known as the *continuity* versus *discontinuity* debate. Rather many studies have been reported that show evidence of language skills in primates (Tomasello and Call, 1997). An interesting observation regarding this issue is the presence of *mirror neurons* (Arbib, 2002) in hominids, cells that fire both when the monkey performs an action in some way or another, or *sees* another monkey perform an action. These cells are also present in humans (in Broca's area). Although there is much speculation regarding the precise function of these neurons, it is conjectured that the play an important role in imitation and learning, social understanding, and mind reading (Origgi and Sperber, 2005).

Although some studies report interesting results of the efforts to teach language to chimpanzees, there are quite a lot of differences between humans and primates on this subject, as Tomasello (2000) pointed out. While it might be true that individual apes have learnt to communicate with humans in a way that presumes at least some intelligent behaviour (e.g., constructing new concepts by combining existing and known concepts), this knowledge stays with the individual. The chimp will normally not try to teach his or her new knowledge to her offspring or other apes, and the newly learnt skills will disappear eventually when the individual dies.

Apart from the question *how* language evolved, another interesting question, discussed by Livingstone (2003), is *why* it evolved, i.e. what the benefits of using spoken language, rather than signs, are. Livingstone discusses grooming (maintaining relationships between individuals), gossip (transferring information about others), second hand information, cheating (gain at the expense of others) and mate and kin selection as advantages for the use of language. However interesting the rationale of language evolution might be, in this thesis I have only investigated the evolution of language as a linguistic ability.

4.2 Language games as a model for language evolution

As pointed out in section 1.6, language game simulations are used as an indirect evidence for language evolution. However abstract these simulations might be, models that are built on language games do possess (at least some) validity, because of their ability to formulate – and test – predictions regarding specific architectures related to language evolution.

In order to be plausible, such models must capture the distinctive features of the architectures one wants to compare. Furthermore, one needs to show that other aspects, that were abstracted away in the model, are not of vital importance, and that the experiments 'scale up', i.e. that the results are independent of the specific configuration of the experiments. Regarding the first aspect, the simulations described in this thesis are built on existing models from the literature (see Vogt & Coumans, 2003, for an overview). The enhancements of the observational game capture the essential features of following and direction attention – contrasted with sharing or checking attention – namely the ability to attend to objects (passively or actively) that *were previously not in the scope of the actors*.

The simulations in this theses were limited with respect to the number of agents, number of language games, number of distinct properties and objects in the environment. Partially this limitation was due to time constraints; other constraints were imposed by the specific implementation using 3APL agents. Vogt and Coumans (2003) found, that increasing the population size from 2 to 20 agents, resulted in only a small decrease in communicative success and coherence rate in the observational and guessing games, and a large decrease in the selfish games. Apart from the results of the selfish game, this is in line with my conclusion that the results are not affected by an increase in population size from 2 to 3.

4.3 Conclusions

The results of the simulations show gradual, rather than dramatic, improvements in the language games when the agents have joint attentional capabilities. In simple cases, feedback can be just as successful as a strategy for lexical development. These results suggest that the ability to engage in joint attentional relations is not the ‘crucial ingredient’ that is necessary for language evolution and the development of a Theory of Mind, as the escalator model in figure 3 suggests. Based on these results, are more plausible model would involve other basic cognitive skills, like imitation, next to joint attention as precursors for language evolution and the development of a Theory of Mind.

The results of the language games in ambiguous environments are more difficult to interpret. Recall, that in ambiguous environments a certain amount of objects was ‘seen’ differently by both agents, thus simulating a sort of deictic interpretation of the object. Although the actual environments consisted of abstract attribute values (like e1 and e5) that differed with a certain probability for agents A and B, the idea can be visualized if we see these attributes as properties like colour and shape. Then, one agent might percept a three-dimensional box as a rectangle, and another agent might see it as a square. If there are other squares and rectangles in the environment, the ability to refer to other objects with similar properties would only add to the confusion of the agents in such situations.

This might explain why the coherence rates of the joint attentional enhancements drops to the level of the non-enhanced observational games when the environment gets ambiguous: the advantage of being able to be more specific about the topic of the language game, becomes a disadvantage when the objects are perceived differently by the agents. Recall that the coherence rate was defined to measure to what extent the agents use the same label for the same concept, and that the weights of *all* label-concept associations that match the agents’ beliefs are increased. Being able to be more specific about the exact topic of the game is not helpful if the agents interpret this topic differently. To put it in an other way: when there is uncertainty about the exact topic of the language game (i.e., there is more than one concept that might be applicable), the ‘error’ of the ambiguous object is ‘spread out’ over more concepts and thus has less impact. These results lend credence to the hypothesis that we need a Theory of Mind in order to cope with deictic relations, false beliefs and opaque contexts: we need to ‘stand in the other one’s shoes’.

The results of the guessing game are intriguing, because the coherence rate actually increases – although not significantly – in more ambiguous environments. This might be explained as follows: since the guessing game is based on coincidental matches in the first stages (when the agents guess a meaning for an utterance), there might be some benefit from ambiguous environments to overcome *false positive* matchings. For example, suppose a game was played in which the first agent referred to a label which matched ‘red’, where the other believed it meant ‘square’ and pointed to a red square. This would lead to a false positive matching, incorrectly increasing the weights for ‘red’ and ‘square’, respectively. If, in a subsequent game, the first agent sees a red square and the other a red rectangle, the *false negative* failure to agree on the square object might compensate for this effect. This is of course a rather speculative explanation of these results, so further research is necessary to resolve this question.

An interesting comparison can be made between the results of this study, and the data of studies involving autistic children. According to Baron-Cohen (1995), these children lack some fundamental joint attentional mechanisms: probably they do have EDD and ID, but lack SAM[5]. With respect to a Theory of Mind, Baron-Cohen, Leslie, and Frith (1985) found that 70% of these children do not pass the False Belief Test at almost 12 years of age. In a more advanced test (Baron-Cohen, 1989) involving nested beliefs, most teenagers with autism fail this test, unlike most 7-year old children without autism. Finally, Bartolucci and Albers (1974) found that autistic children had great difficulties with interpreting sentences that involved deictic structures.

4.4 Further research

In this thesis, I modelled the language games using the 3APL programming environment, which constrained the number and length of simulations that could be run. Existing packages for language games, such as THSim[6], that are more suitable for large scale-experiments, could be enhanced with the joint attentional enhancements discussed in this thesis in order to obtain more results. For example, it would be interesting to see how the enhanced games performed after 10,000 games. Especially the results of the guessing game in ambiguous conditions suggest further research to clarify these unexpected findings.

Furthermore, the concept of *ambiguity* in the context of language games could be further explored. In the language games described in this thesis, ambiguity was defined as simply ‘perceiving objects differently’, focusing on ambiguity that originated from the assumed different perspectives of the agents. On the other hand, a more intuitive meaning of ambiguity might be that objects are under-defined or under-specified, e.g. the utterance ‘Look at that tree!’ in a forest where it could be difficult to decide to which exact tree one is pointing.
Other enhancements to the language games might deal with mechanisms that are traditionally associated with learning, language and ToM, such as imitation and pretend play.

[5] SAM: Shared Attention Mechanism; ID: Intentionality Detector, EDD: Eye Direction Detector
[6] http://www.ling.ed.ac.uk/~paulv/thsim.html

References

Arbib, M.A. (2002) The mirror system, imitation and the evolution of language. In C. Nehaniv and K. Dautenhahn (Eds.), *Imitation in Animals and Artifacts*, Cambridge, MA: MIT Press.

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21, 37–46.

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry*, 30, 285–298.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind.* Cambridge, MA: MIT Press.

Bartolucci, G., & Albers, R.J. (1974). Deictic categories in the language of autistic children, *Journal of Autism and Developmental Disorders*, 4(2), 131–141.

Carey, S. (1978). The child as word-learner. In M. Halle, J. Bresnan, and G.A. Miller (Eds.), *Linguistic theory and psychological reality*. Cambridge, MA: MIT Press.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 63(4, Serial No. 255).

Christiansen, M. & Kirby, S. (2003). *Language Evolution*. Oxford University Press.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, 15, 481–498.

Chomsky, N.A. (1968). *Language and Mind*. New York: Harcourt Brace.

A.G. Clark, S. Glanowski, R. Nielsen, P.D. Thomas, A. Kejariwal, M.A. Todd, D.M. Tanenbaum, D. Civello, F. Lu, B. Murphy, S. Ferriera, G. Wang, X. Zheng, T.J. White, J.J. Sninsky, M.D. Adams, & M. Cargill. (2003). Inferring Nonneutral Evolution from Human-Chimp-Mouse Orthologous Gene Trios. *Science*, 12 December 2003, 1960–1963.

Davis, K. (1947). Final note on a case of extreme social isolation. *American Journal of Sociology*, 52, 432–437.

Hurford, J.R. (2004). Language beyond our grasp: What mirror neurons can, and cannot do, for language evolution. In D. Kimbrough Oller, U. Griebel & K. Plunkett (Eds.) *Evolution of Communication Systems: A Comparative Approach* (pp. 297–313). The Vienna Series in Theoretical Biology. Cambridge, MA: MIT Press.

Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Kohler, W. (1929). *Gestalt psychology*. New York: H. Liveright.

Lieven, E.V.M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway and B. J. Richards (Eds.), *Input and interaction in language acquisition* (p. 56–73). Cambridge: Cambridge University Press.

Livingstone, D. (2003). *Computer Models Of The Evolution Of Language And Languages*. PhD-thesis, University of Paisley.

Malle, B.F. (2002). The relation between language and theory of mind in development and evolution. In T. Givón & B. F. Malle (Eds.), *The evolution of language out of pre-language* (p. 265–284). Amsterdam: Benjamins.

MacWhinney, B. (1998). Models of the Emergence of Language. *Annual Review of Psychology,* 49, 199–227

Pan, B.A. (2005) Semantic Development: Learning the Meaning of Words. In Gleason (ed.), The development of language, edition 2005.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, (13), 707–784.

Oliphant, M. (1997). The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior, 7,(3-4*), 371–384.

Origgi, G., & Sperber, D. (2005) What do mirror neurons mean? Virtual conference, website http://www.interdisciplines.org/mirror/

Reboul, A. (2004). Evolution of Language from Theory of Mind or Coevolution of Language and Theory of Mind? In: Issues in Coevolution of Language and Theory of Mind, virtual conference, website http://www.interdisciplines.org/coevolution/

Robinson, E.J., & Apperlyb, I.A. (2001). Children’s difficulties with partial representations in ambiguous messages and referentially opaque contexts. *Cognitive Development*, 16, 595–615.

Roessler, J. & Eilan, N. (2003). *Agency and Self-Awareness*. Issues in Philosophy and Psychology. Clarendon press, Oxford.

Smith, A.D.M. (2001). Establishing Communication Systems without Explicit Meaning Transmission. In J. Kelemen and P. Sosik (Eds.), *Proceedings of the 6th European Conference on Artificial Life, ECAL 2001* (p. 381–390). Springer-Verlag.

Steels, L. (1999a) *The Talking Heads Experiment. Volume 1. Words and Meanings*. Laboratorium, Antwerpen.

Steels, L. (1999b). The Puzzle of Evolution. *Kognitionswissenschaft*, 8(4), 143–150.

Steels, L. (2001). Language games for autonomous robots. *IEEE Intelligent Systems,* 16(5), 16–22.

Steels, L., & Kaplan, F. (2002). Bootstrapping grounded word semantics. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: formal and computational models*. Cambridge University Press.

Sterelny, K. (2000). Primate Mind Readers. In C. Heyes and L. Huber (Eds.). *The evolution of cognition*. The Vienna series in theoretical biology.

Tomasello, M. (1995). Language is Not an Instinct. *Cognitive Development*, 10, 131–156.

Tomasello, M., & Call, J. (1997). *Primate Cognition*. Oxford University Press.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press.

Tomasello, M. (2000). Two hypotheses about primate cognition. In C. Heyes and L. Huber (Eds.). *The evolution of cognition*. The Vienna series in theoretical biology.

Vogt, P., and Coumans, H. (2003). Investigating social interaction strategies for bootstrapping lexicon development. *Journal of Artificial Societies and Social Simulation* 6(1).

Wimmer, H. and Perner, J. (1983). Beliefs about beliefs: Representation and constraining function in wrong beliefs in young children’s understanding of deception. *Cognition*, 13, 103–128

Appendices

1 All results of language games

In this appendix the results of all language games are given, in all conditions (baseline, three properties, three agents, slightly ambiguous, moderately ambiguous, iterated learning). For all conditions the data for all six game types (guessing game, selfish game, observational game, and the three joint attentional enhancements). Each table consists of 10 simulation runs with different random seeds for agent A, B and (optionally) C. The coherence rates per concept (e1-e6) and their means were recorded.

Baseline condition (2 agents, 2 properties, unambiguous environment)

	GG	SG	OG	JA1	JA2	JA3
1	0.3033	0.0664	0.2737	0.2672	0.4315	0.3979
2	0.0844	0.1257	0.1480	0.3482	0.2329	0.4327
3	0.5569	0.0876	0.2790	0.3094	0.4127	0.3165
4	0.2492	0.0430	0.0922	0.3173	0.3501	0.3587
5	0.3589	0.0337	0.1296	0.2290	0.2776	0.4437
6	0.3516	0.0777	0.2798	0.3231	0.3396	0.5394
7	0.2596	0.0302	0.1203	0.3636	0.3638	0.4414
8	0.2675	0.0696	0.2553	0.3274	0.3306	0.4778
9	0.5764	0.1142	0.1994	0.2597	0.2993	0.4391
10	0.2492	0.0712	0.2901	0.3775	0.2497	0.2711
Mean	**0.3257**	**0.0719**	**0.2067**	**0.3122**	**0.3288**	**0.4118**
St.D	**0.1477**	**0.1227**	**0.0777**	**0.0475**	**0.0653**	**0.0786**

Three properties

	GG	SG	OG	JA1	JA2	JA3
1	0.1491	0.0307	0.1045	0.1257	0.1597	0.1662
2	0.1194	0.0348	0.0750	0.1704	0.1216	0.1239
3	0.1806	0.0363	0.0443	0.2197	0.1620	0.2218
4	0.2018	0.0067	0.0373	0.1731	0.1425	0.1896
5	0.1655	0.0323	0.0466	0.1607	0.1391	0.1502
6	0.0061	0.0334	0.0440	0.1571	0.1502	0.1574
7	0.1733	0.0295	0.1149	0.1325	0.1330	0.1510
8	0.1566	0.0206	0.0435	0.1599	0.1893	0.2226
9	0.0904	0.0109	0.1290	0.1978	0.2242	0.2418
10	0.0778	0.0159	0.0896	0.1610	0.2528	0.1844
Mean	**0.1321**	**0.0251**	**0.0729**	**0.1658**	**0.1674**	**0.1809**
St.D	**0.0593**	**0.0107**	**0.0344**	**0.0276**	**0.0422**	**0.0381**

Three agents

	GG	SG	OG	JA1	JA2	JA3
1	0.3033	0.0760	0.2729	0.2335	0.4318	0.4574
2	0.0844	0.0702	0.2361	0.3491	0.2131	0.3503
3	0.5569	0.0662	0.2830	0.3990	0.3533	0.3967
4	0.2492	0.1180	0.2324	0.1792	0.2659	0.2451
5	0.3589	0.0825	0.1725	0.3210	0.3319	0.4724
6	0.3516	0.0301	0.2455	0.2795	0.3686	0.3603
7	0.2596	0.0305	0.0956	0.3255	0.4301	0.4403
8	0.2675	0.1097	0.2479	0.1952	0.3707	0.3451
9	0.5764	0.0381	0.1086	0.3424	0.2702	0.3508
10	0.2492	0.0523	0.2677	0.2576	0.4237	0.4423
Mean	**0.3257**	**0.0674**	**0.2162**	**0.2882**	**0.3459**	**0.3861**
St.D	**0.1477**	**0.0307**	**0.0674**	**0.0714**	**0.0758**	**0.0695**

Slightly ambiguous

	GG	SG	OG	JA1	JA2	JA3
1	0.3033	0.0677	0.2737	0.2741	0.4214	0.3959
2	0.0844	0.1257	0.1480	0.3519	0.2346	0.4050
3	0.5569	0.0377	0.2888	0.2918	0.3901	0.2916
4	0.2492	0.0430	0.0958	0.3124	0.3325	0.3463
5	0.3589	0.0343	0.1319	0.2404	0.2941	0.4525
6	0.3516	0.0777	0.2798	0.3337	0.3359	0.3767
7	0.2596	0.0299	0.1116	0.3380	0.3385	0.4009
8	0.2675	0.0676	0.2323	0.3302	0.3220	0.4385
9	0.5764	0.1204	0.2002	0.2632	0.2922	0.4085
10	0.2492	0.0720	0.2901	0.3676	0.2451	0.2039
Mean	**0.3257**	**0.0676**	**0.2052**	**0.3103**	**0.3206**	**0.3720**
St.D	**0.1308**	**0.0339**	**0.0779**	**0.0415**	**0.0580**	**0.0746**

Moderately ambiguous

	GG	SG	OG	JA1	JA2	JA3
1	0.4959	0.0696	0.2757	0.1665	0.2770	0.2357
2	0.0844	0.1206	0.1694	0.0849	0.3198	0.3647
3	0.5569	0.0331	0.2800	0.2339	0.2286	0.2010
4	0.2492	0.0519	0.1277	0.1741	0.1335	0.1318
5	0.4843	0.0352	0.1333	0.2411	0.2229	0.2232
6	0.3763	0.0782	0.2758	0.2227	0.1961	0.2462
7	0.1175	0.0401	0.0275	0.3079	0.3125	0.2963
8	0.2808	0.0728	0.2748	0.3116	0.2590	0.2118
9	0.5618	0.1201	0.2018	0.1480	0.3425	0.1621
10	0.2176	0.0712	0.2795	0.2844	0.2298	0.1799
Mean	**0.3425**	**0.0693**	**0.2046**	**0.2175**	**0.2522**	**0.2253**
St.D	**0.1777**	**0.0315**	**0.0882**	**0.0740**	**0.0666**	**0.0672**

Iterated learning

	GG	SG	OG	JA1	JA2	JA3
1	0,1050	0,1983	0,4150	0,5533	0,7000	0,7000
2	0,1050	0,2200	0,5050	0,5683	0,6067	0,6233
3	0,3783	0,1983	0,4150	0,5533	0,5400	0,6133
4	0,4383	0,2233	0,4567	0,5950	0,5650	0,6450
5	0,2500	0,2167	0,4917	0,6100	0,6317	0,6733
6	0,3500	0,2217	0,4300	0,5483	0,6500	0,6500
7	0,3650	0,1583	0,4300	0,5667	0,5217	0,6333
8	0,3450	0,2233	0,4850	0,6300	0,7067	0,7067
9	0,5083	0,2217	0,4300	0,5483	0,5217	0,5750
10	0,1467	0,1667	0,4567	0,5333	0,6767	0,6783
Mean	**0,2992**	**0,2048**	**0,4515**	**0,5707**	**0,6120**	**0,6498**
St.D	**0,1412**	**0,0243**	**0,0328**	**0,0311**	**0,0688**	**0,0409**

2 Pseudocode of the language games

2.1 guessing game

Pseudocode guessing game agent

```
/* In this main loop we load an environment, play a game, update statistics and make
   sure the other agent is ready as well */

languageGame(int Max, agent Agent)
BEGIN
       WHILE Game < Max DO
       BEGIN
              Game = Game + 1
              getEnvironment(Game, Agent)
              playGame(Game)
              updateStatistics()
              removeGarbage()
              synchronize()
       END
END

/* A and B play alternatively as Talker and Hearer. The code for this part differs
   ofcourse for both agents. */

playGame(int Game)
BEGIN
       IF odd(Game) THEN playTalkerGame() ELSE playHearerGame()      [A]
       IF even(Game) THEN playTalkerGame() ELSE playHearerGame()     [B]
END

/* The talker selects a random concept, and finds the label with the highest weight.
   the talker then communicates the label and waits for an incoming object. If this
   object indeed has the concept the talker selected, the weight for this specific
   combination of concept and label is updated and the weight of conflicting
combinations
   is decreased. Success or failure is reported to the hearer. */

playTalkerGame()
BEGIN
       Concept = selectRandomConcept()
       IF hasKnownLabel(Concept) THEN
       BEGIN
              Label = findHighestLabel(Concept)
              send(OtherAgent, Label)
       END
       ELSE BEGIN
              generateLabel()
              send(OtherAgent, Label)
       END

       WaitUntilReceived(Object)

       IF hasConceptWithLabel(Object, Concept, Label) THEN
       BEGIN
              addWeight(Concept, Label)
              subWeight(Concept, OtherLabels)
              subWeight(OtherConcepts, Label)
              send(OtherAgent, ack(Object, Label))
       END
       ELSE BEGIN
              subWeight(Concept, Label)
              send(OtherAgent, nak(Object, Label))
       END
END

/* The hearer waits until it receives a label. It finds a the concept with the
   highest wait value for this label – or asserts a random concept – and looks in
```

```
   the environment for a matching object. The weights are adjusted based on feedback
*/

playHearerGame()
BEGIN
       WaitUntilReceived(Label)
       IF unknown(Label) THEN
       BEGIN
              Concept = selectRandomConcept()
              assert(Concept, Label)
       END
       ELSE BEGIN
              getMatchingConcept(Label)
              findMatchingObject(Concept);

              send(OtherAgent, Object)
              waitUntilReceived(Feedback)

              IF Feedback = ack(Object, Label) THEN
              BEGIN
                     addWeight(Concept, Label)
                     subWeight(Concept, OtherLabels)
                     subWeight(OtherConcepts, Label)
              END
              ELSE IF Feedback = nak(Object, Label) THEN
                     subWeight(Concept, Label)
       END
END
```

2.2 observational game

Pseudocode observational game agent

```
/* In this main loop we load an environment, play a game, update statistics and make
   sure the other agent is ready as well */

languageGame(int Max, agent Agent)
BEGIN
       WHILE Game < Max DO
       BEGIN
              Game = Game + 1
              getEnvironment(Game, Agent)
              playGame(Game)
              updateStatistics()
              removeGarbage()
              synchronize()
       END
END

playGame(int Game)
BEGIN
       IF odd(Game) THEN playTalkerGame() ELSE playHearerGame()      [A]
       IF even(Game) THEN playTalkerGame() ELSE playHearerGame()     [B]
END

/* The talker selects a random concept, and finds an object in the environment that
has
   this concept and communicates the object. It then waits for a reaction which might
be
   a label or a request. If the hearer has sufficient capabilities, it can ask for a
   similar object or show an alternative object. The talker can respond apropriately
if
   it has sufficient capabilities. After a label is received, the corresponding
weights
   are updated. Note that no feedback is given to the hearer. */

playTalkerGame()
BEGIN
       Concept = selectRandomConcept()
       Object = selectObjectWithConcept(Concept)
       Send(OtherAgent, Object)

       WHILE (Reaction <> SENT_LABEL) DO
       BEGIN
              WaitUntilReceived(Reaction)

              IF Reaction = QUERY_OTHER_OBJECT AND Stadium = DIRECT THEN
              BEGIN
                     AlternateObject =
                            selectOtherObjectWithConcept(Object, Concept)
                     Send(OtherAgent, AlternateObject)
              END
              IF Reaction = (ASK_OTHER_OBJECT, AlternateObject) AND Stadium = FOLLOW
       THEN
              BEGIN
                     IF hasConcept(AlternateObject, Concept) THEN
                            Send(OtherAgent, ACK)
                     ELSE
                            Send(OtherAgent, NAK)
              END
              IF Reaction = (SENT_LABEL, Label) THEN
              BEGIN
                     addWeight(Concept, Label)
                     subWeight(Concept, OtherLabels)
                     subWeight(OtherConcepts, Label)
              END
       END
```

```
END

/* The hearer waits until it has received an object and tries to further specify the
   set of concepts that are applicable, using joint attention. */

playHearerGame()
BEGIN
       WaitUntilReceived(Object)

       IF Stadium = NO_JOINT_ATT THEN Object = selectRandomObject()
       ELSE Concepts = SET OF getConcepts(Object)

       /* If we can follow direction, ask the talker to show us an alternative
          object that also has the concept that the talker had in mind. */

       IF Stadium = FOLLOW THEN
       BEGIN
              Send(OtherAgent, QueryOtherObject)
              WaitUntilReceived(OtherObject)
              AlternateConcepts = SET OF getConcepts(OtherObject)
              Concepts = intersection(Concepts, AlternateConcepts)
       END

       /* If we can direct direction, show the talker an alternative object
          and ask whether it also has the concept that the talker had in mind. */

       IF Stadium = DIRECT THEN
       BEGIN
              OtherObject = selectOtherObject()
              AlternateConcepts = SET OF getConcepts(OtherObject)
              Send(OtherAgent, OtherObject)
              WaitUntilReceived(Feedback)
              IF Feedback = ACK THEN
                     Concepts = intersection(Concepts, AlternateConcepts)
              ELSE IF Feedback = NAK THEN
                     Concepts = disjoint(Concepts, AlternateConcepts)
       END

       /* Find and communicate the highest applicable label.

       IF hasKnownLabel(Concepts) THEN
              Label = findHighestLabel(Concepts)
       ELSE
              Label = generateLabel()
       send(OtherAgent, Label)

       FORALL Concept IN Concepts  DO
       BEGIN
              addWeight(Concept, Label)
              subWeight(Concept, OtherLabels)
       END
       FORALL Concept NOT IN Concepts DO
       BEGIN
              subWeight(Concept, Label)
       END
END
```

3. Settings and parameters compared with De Vogt and Coumans (2003)

Parameter/Setting	De Vogt and Coumans	This study
Number of simulation runs	10	10
Number of language games	50000	100
Context size	5	4
Number of agents	2-20	2-3
Number of meanings	100	3x2 to 3x3
Initial Association Score	$\sigma = 0.01$	$\sigma = 0.01$
Learning Rate	$\eta = 0.9$	$\eta = 0.9$
Series	3	6

4. Construction of the environments

The environment of the agents was constructed such, that four objects were shown with two properties, which each could have distinct values. The four objects were chosen at random from a set of all possible combinations. It was possible to have more than one copy of the same object.

A similar procedure was followed for the games where all object had three possible properties. In ambiguous environments, there was a 10% probability in the slightly ambigious condition and a 50% probability in the moderately ambiguous condition that a concept was not the same for agent A and B, i.e. one of the concepts in the environment in the environment for A was changed.

5 Protocols of the False Belief and Opaque Context tests

5.1 False Belief Test

The participant is shown a scene with a box, a basket, and two persons: Sally and Anne. Sally puts a toy in the basket and leaves the scene. Then Anne removes the toy from the basket and puts it in the box, visible for the participant but not for Sally. Then Sally returns, and the participant is asked where she will look for the toy.

5.2 Opaque Context Test

In the Opaque Context Test, participants are presented scenarios and have to answer questions that tested whether they understood word substitution in transparent (a-sentences) and opaque (b- sentences) questions. Some examples of these scenarios and questions could be:

One scenario involved telling children a story about a little boy and girl who went to the doctor's office to get a bandage for the girl (Anna). The boy stood next to the doctor to watch while she put a bandage on the little girl. Unbeknownst to the boy (Mark), the doctor was Anna's mom.

a) Was Mark standing next to Anna’s mom?
b) Did Mark knew that he did so?

In another scenario, a character was both "Sue's dad" and a "police officer." A little boy in the story, Mark, who knew this character only as a police officer, handed over some dropped keys.

a) Did Mark gave Sue's dad the keys?
b) Did Mark knew that he gave Sue's dad the keys?