Luc Steels

Language Re-Entrance and the 'Inner Voice'

Abstract: *As soon as we stop talking aloud, we seem to experience a kind of 'inner voice', a steady stream of verbal fragments expressing ongoing thoughts. What kind of information processing structures are required to explain such a phenomenon? Why would an 'inner voice' be useful? How could it have arisen? This paper explores these questions and reports briefly some computational experiments to help elucidate them.*

Keywords: consciousness, origins of language, inner voice, robotic models.

I: Introduction

Although consciousness is sometimes believed to involve a single explanatory principle (and would therefore be something like gravity, see Penrose,1990), it has so far proven more fruitful to regard it as a combination of many different aspects. These include: unifying different views of reality experienced through different sensory modalities into a coherent picture, focusing attention, feeling emotions, acting coherently by combining top-down goals with bottom-up reactive behaviour, forming thoughts and expressing them in words, building and maintaining a self-model, intuiting sudden insights (the 'Aha' experience), making and justifying ethical decisions, and having a subjective phenomenal experience of the self.

The main reason for analysing consciousness this way comes from convincing evidence that there is a functional specialization of brain areas and regions. Visual awareness is correlated to coalitions of active neurons in the neocortex (Crick and Koch, 1995), speech production and perception shows strong activations of Broca's and Wernicke's area, emotion heavily exercises the amygdala and the hormonal system (LeDoux, 1998). Neuroscience has concentrated enormous efforts in recent decades in order to map these functional specializations,

Correspondence: Luc Steels, Free University of Brussels, VUB AI Laboratory, 10G-725, Pleinlaan 2, B-1050 Brussels, Belgium. *Email: steels@arti.vub.ac.be*

i.e. to find the neural correlates of various mental capacities, including those associated with consciousness (Edelman, 1989).

However knowing where some activity relevant to a certain mental function takes place is not yet a satisfactory explanation. Food is digested in the stomach and the liver. Fine. But we want to know *how* the digestion takes place, *why* food needs to be digested, and *why* the digestive products are necessary for survival. Moreover mental functions might be less localized than they are thought to be (Rosenfield, 1992), just as food digestion actually involves the whole body and its behaviour.

As Dennett has suggested (Dennett, 1991), building very detailed operational models explicating the information processing necessary and sufficient for certain mental functions is one way to address the how and why of some aspects of conscious processes, and AI researchers have indeed been doing this, sometimes with considerable success. The work on modelling (shared) attention is a good example of this approach.

Children are already capable of establishing shared attention and turn-taking after a few months of life, and it forms a necessary basis for the social learning of language and meaning (Tomasello, 1999). The many components involved in attention sharing are beginning to be understood and several AI researchers have recently built integrated robotic systems that combine a large set of these mechanisms to achieve attentional behaviour (Scassellati, 1998). They include the ability to pick from a visual image stream the objects that are salient in a particular context, such as faces, separate out sound sources and identify visually the object responsible for the sound, keep tracking objects as they move, and observe what a person is attending to by detecting pointing gestures and the orientation of the face. These mechanisms are being integrated on a new generation of humanoid robots (such as KISMET [Breazeal, 1998] or the Sony SDR robot [Kuroki *et al*., 2001]), yielding impressive human-like performance.

Are the robots employing these mechanisms conscious? Most people would say no — even though it is quite extraordinary to interact with them. But these developments show that whatever consciousness 'really' is, some of the behaviours often associated with having consciousness can be unravelled, and their information processing foundations understood.

This paper adopts the same strategy for studying another aspect of human behaviour associated with consciousness. Every normal human being, including deaf people, report hearing almost constantly a silent voice when they are not overtly speaking. This inner voice jumps from one topic to the other. It seldom produces very well formed sentences but rather fragmented bits and pieces of thought, as captured by Joyce's *Ulysses*. The inner voice is there as soon as we wake up and particularly when we are trying to write or form new thoughts. People fluent in several languages report that their inner voice occasionally speaks another language. The inner voice is closely linked to the self, indeed we have the feeling that it is the 'I' which is talking, like in a first person narrative. It is also possible to hear the voice of somebody else, such as when we imagine a dialogue or hear people speak in a dream.

The questions raised in this paper are going to be the following: Can we unravel the information processing structures achieving this mental functioning? How exactly might such an inner voice be produced? And why would an inner voice have arisen in the course of evolution? What are additional benefits once such an inner voice is there?

In summary, my argument is the following:

(1) Achieving an inner voice is technically not so difficult. The language system must be turned onto itself. Output from the subsystem that produces speech must be fed back into the subsystem that understands it. Many psycholinguistic models going back to the nineteenth century show this feature and some of our recent computer simulations do so as well.

(2) The why question also has a straightforward answer. The re-entrant mapping of the language system is a means for bootstrapping language towards higher complexity, indeed I believe it is a prerequisite. Many psycholinguists have argued that natural language arose as an externalization of the internal language of thought (Fodor, 1975). I will argue just the opposite. Internal language is a side effect of being able to use and learn external language.

(3) The 'inner voice mechanism' is not only useful in language construction and language learning. It can also be used to come to terms with our experiences. It enables a simulation of the thought processes of others and thus provides the foundation for complex social behaviour.

The remainder of this paper discusses these points in more detail. I will report on some computational and robotic experiments, trying to avoid as much as possible over-technical aspects.

II: Language Communication

1. Language as a complex adaptive system

We need a theory of language which differs from the one common in contemporary Chomskyan linguistics. Generative grammar has been conceived as a descriptive model of the syntactic regularities of languages, whereas we need a model of how language production and interpretation works in concrete settings. There are three characteristics of language which I want to emphasize:

1. Language arose for communication. We speak primarily because we want to influence the thoughts and behaviours of somebody else (Austin, 1975) This does not mean that there are no derived uses of language, as in religious ceremonies. But denying the communicative function of language makes it incomprehensible why we speak and also why languages are the way they are. So, the generative source of speech does not lie in syntax, as in the Chomskyan model, but in subsystems that conceptualize and select meaning.

2. Natural language does not fit with the simplest form of the Shannon model of communication. In a Shannon model, there is a sender and receiver, a

message, and a message content. Production and interpretation is considered to be equal to a coding and decoding process. The content is assumed to be completely present in the message. In contrast, communication in natural language assumes an intelligent receiver, i.e. an agent that shares the same context and can fill in a large part of what the sender wants to transmit (Sperber and Wilson, 1986). This is why the computer processing of natural language is not really possible, unless a computer or robot has the same inferential powers and shares the same environment and contexts as humans.

The inferential aspect of natural language communication is its main power but also its principal weakness. It is more powerful than a Shannon system because much more complex information can be communicated with fewer means. But it is weaker because much more processing needs to be done both by the speaker (who has to guess what the hearer might already know or be able to infer) and by the listener (who has to guess what the speaker might mean). Moreover there is a higher risk of misunderstanding if speaker and hearer do not sufficiently share a common ground.

3. Language is in constant evolution. Every verbal interaction is a creative one and lexical and grammatical regularity is emergent and transient (Hopper, 1991). Speakers and hearers create new conventions for the occasion (Clark and Brennan, 1991) and these might then propagate to the rest of the community. Innovations include new words, new meanings for existing words, new grammatical constructions, and new patterns of dialogue. These micro-changes in language are perhaps not visible to language participants but they are blatantly obvious when one examines language usage even over a very short stretch of time. A person's language is always in flux and any model of language processing must include components to explain how participants in a conversation adapt the language to suit their needs.

2. The semiotic cycle

Obviously the formulation of an utterance and its interpretation involve highly complex processes, particularly when the communication is grounded in a concrete situation and therefore implies perception and action. Most of these processes are completely unconscious. We have no direct access to the intermediary structures nor to the grammatical knowledge we all possess. And it is best that way. Language processing is so complex that it would not be possible to keep every intermediary step in mind, particularly because processing must be fast enough to respond fluently in dialogue.

The main processes involved in language comprehension and production are displayed in the semiotic cycle (see figure 1). They take place against the background of turn-taking and attention sharing behaviours and are of course strongly intermingled and influencing each other.

(1) *Grounding*. The first set of processes carried out by the speaker must maintain a connection between the internal factual memory and the states and actions in the world the speaker wants to talk about. They

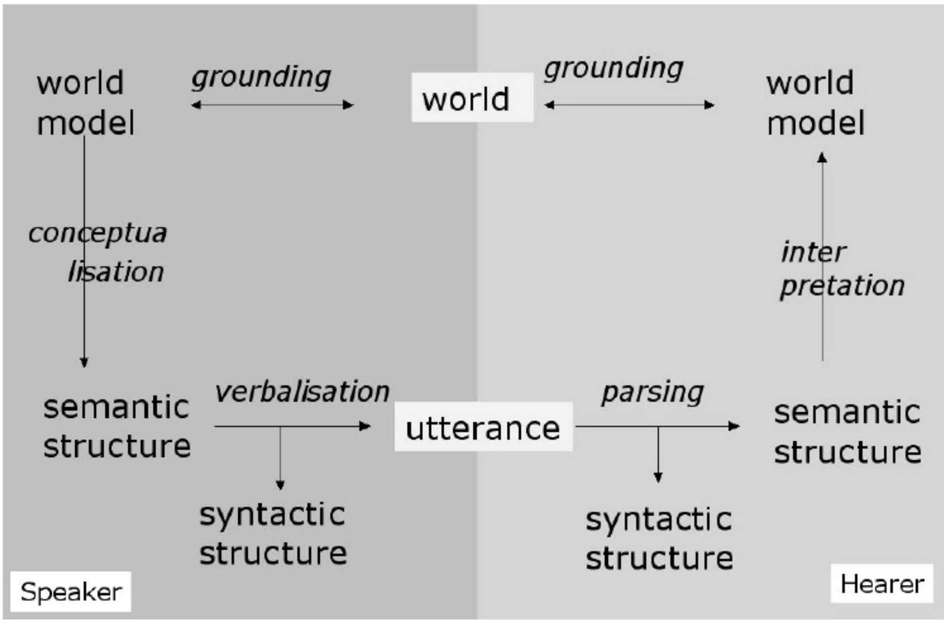

Figure 1. The semiotic cycle

Left: processes carried out by the speaker. Right: processes carried out by the hearer.

include segmentation, feature extraction, object recognition, event classification, object tracking, object manipulation, etc.

(2) *Conceptualisation*. The second set of processes must select what needs to be said and conceptualize the world in a way that it can be translated to natural language (Talmy, 2000). This requires inference and possibly additional perception or action. For example, if we say ‘the car is in front of the tree’, we have conceptualized the tree as having a front which is directed towards us, and the car as being in between ourselves and this front. This conceptualisation is language dependent. In some African languages, the front of the tree is in the same direction as the face of somebody looking at a tree, so ‘the car in front of the tree’ now means behind the tree (Heine, 1997).

(3) *Verbalisation*. This is the set of processes that take a semantic structure and render it through a series of mappings into a surface form. This is an enormously complex process, bringing to bear a vast array of lexical, grammatical, morphological, phonological and phonetic conventions, which are obviously language-dependent. The final outcome of verbalisation is transmitted as an utterance to the hearer.

(4) *Parsing*. The hearer must reconstruct a semantic structure from the utterance that is transmitted. This process is again highly elaborate due to the complexity of natural language and the presence of ambiguities. The steps in parsing cannot simply be chained linearly but must operate as a constraint propagation process.

(5) *Interpretation*. The hearer must confront the semantic structure resulting from the parsing process with his factual memory of the world and understanding of the dialogue context in order to find a correct interpretation, i.e. an interpretation of the utterance that fits with his expectations and observations.
(6) *Grounding*. The hearer must therefore also maintain a connection through perception and action between his internal factual memory and the states of the world, including the states of the speaker.

At present we can construct computational models of all aspects of this cycle (at least in a simplified form), and for our experiments on embodied communication with autonomous robots this is what we do (Steels, 2001a,b; Steels and Kaplan, 2001). The semiotic cycle is embedded in language games, in which speaker and hearer take turns trying to achieve a particular goal (such as drawing attention to an object in the environment through verbal means).

3. Inventing and learning

The semiotic cycle captures the routine application of existing language conventions, but, as argued earlier, this routine application is hardly ever possible. New situations and new meanings that have never been expressed before come up all the time and so language users are constantly expanding and learning new language. This requires two additional processes:

1. *Language invention*. It is unavoidable that situations arise where the speaker does not know an adequate well-entrenched convention for expressing certain meanings. Either the meanings have never been expressed before (think about all the operations that we can now do through Internet which would be unthinkable fifty years ago), or their conventional expression might give rise to misunderstandings in the present context. Moreover the speaker may occasionally want to avoid worn out phrases to better capture the attention of the listener. So a set of processes is needed that implements strategies to invent new language conventions on the spot. The simplest example of a language invention strategy is to chose a new word to express a predicate. Usually the new word is chosen by analogy with an existing one. For example, a computer pointing device is called a mouse by analogy with the shape of a mouse, or clicking on a hyperlink is called surfing by analogy with catching waves. Invention also happens for grammatical constructions, even though it is perhaps less visible. For example, if no satisfying expression exists to express a certain temporal relation, another word or another grammatical construction might be adopted for this purpose. This is how the auxiliary ‘will’ or the construction ‘is going to’ became gradually adopted in English for the grammatical expression of future (Traugott and Heine, 1991).

2. *Language learning*. A language listener must be prepared to learn new conventions at any moment. As argued earlier, a large part of the meaning of an utterance is implicit and must be actively reconstructed. This is only possible because he shares the same situation as well as the broader context with the speaker, makes similar inferences, and conceptualizes the world in a similar way. These

inferential capabilities make it possible for the listener to learn new bits of language. Suppose someone says (while sitting in front of a computer), ‘Can you hand me the mouse’, and the listener does not know what a mouse is. If the speaker then points to or grabs the mouse herself, the listener can easily infer that this is what is called a mouse. The analogy between a mouse and the computer’s pointing device is very suggestive and therefore not so hard to guess or to remember afterwards.

When a population of speakers and listeners applies language invention and language learning on a large scale and in a cumulative fashion, a communication system progressively emerges. In a series of computer and robotic simulations, we have been experimenting with such processes and consistently seen this phenomenon (Steels *et al*., 2002). The artificial agents effectively negotiate language conventions. They monitor whether a particular meaning-form association has been successful in communication by keeping a score. Agents should prefer the meaning-form associations that had the highest score in the past because this gives the best chance for success in the future. This leads to the spontaneous self-organization of a shared set of conventions due a positive feedback between use and success: The score of a convention which contributed to successful communication increases for the agents that use it, hence they will use this convention even more. Agents that did not know the convention will hear it and eventually all agents will have adopted the same convention.

We have done a number of experiments to simulate both the processes in the semiotic cycle and the invention and learning processes outlined in this subsection. These are part of a more encompassing research programme trying to build computational and robotic models for the emergence of grammar. In our experiments, two robotic heads watch scenes, such as the one shown in figure 2 (from Steels and Baillie, 2002) in which a hand picks up a red object. Then they play a language game. For example, a guessing game in which the speaker chooses an object in the scene and then seeks a verbal description so that the hearer can guess what object was chosen, or a description game in which the speaker chooses an event or a series of events that took place, conceptualizes a description of these events, and formulates an utterance. The hearer produces a sign of approval when comprehending the utterance and agrees that it describes the scene. An experiment typically focuses on a certain aspect of language, such as evolving a lexicon, or some area of grammar.

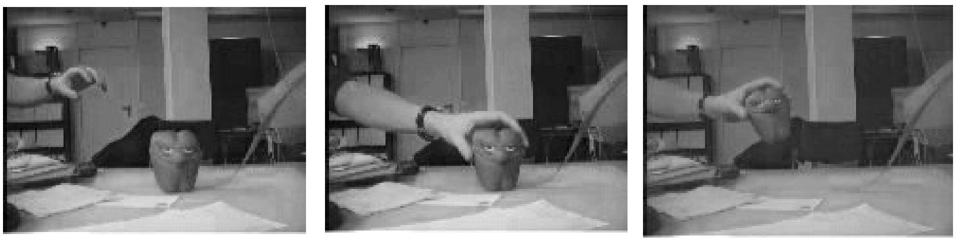

Figure 2. Our data in the emergence of language typically involve two robotic ‘heads’ which observe situations like the one shown here. The robots must use verbal communication to identify objects in the scene or describe events.

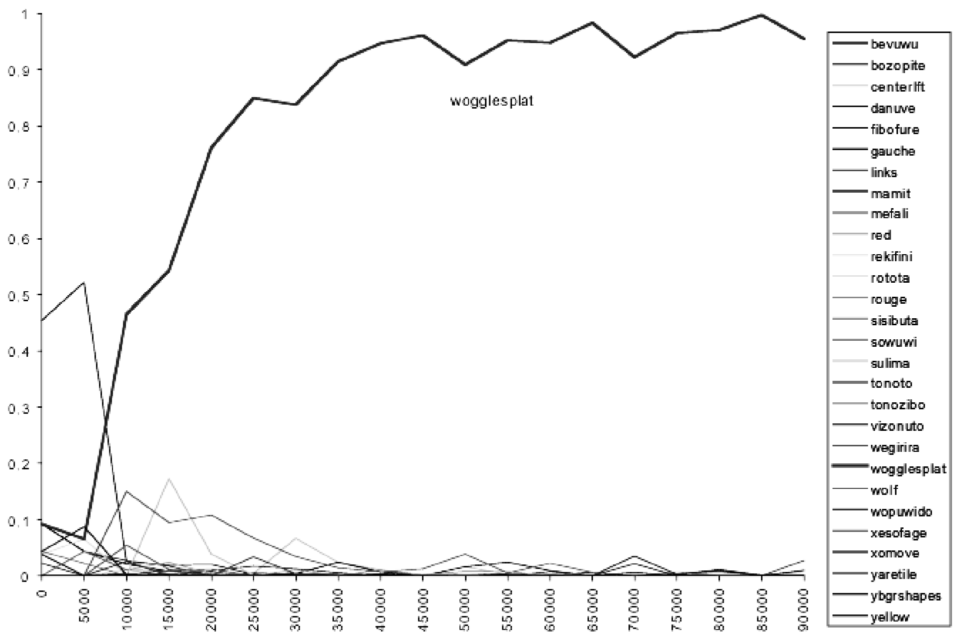

Figure 3.

A meaning-form diagram which graphs for a specific meaning all the possible forms and their frequency of use. A winner-take-all situation is clearly observed. The x-axis shows language games and y-axis the score of forms.

Figure 3 (from Steels *et al.*, 2002) is a typical outcome of an experiment for evolving a lexicon in a steadily growing population reaching 1500 agents towards the end. It shows for a particular meaning, the frequency with which agents use a certain word. The positive feedback between use and success ensures that a winner take all effect takes place: After some time a single word becomes dominant and then propagates in the population, even to new agents that enter the population at regular time intervals.

III: Re-Entrance

1. Rewiring the language system

Now I come to the core issue of this paper. Quite quickly, we were successful in achieving the emergence of a lexicon in our simulations, but the emergence of grammar has turned out to be much more difficult. A breakthrough came as soon as we started to explore the idea that a speaker might apply his own language system (parser and interpreter) to his own utterances, either before transmitting them to the hearer or after observing incomprehension or communicative failure from the part of the listener (figure 4).

It may seem that listening again to a sentence that you produced yourself would not tell you anything new. But that is not the case. The language

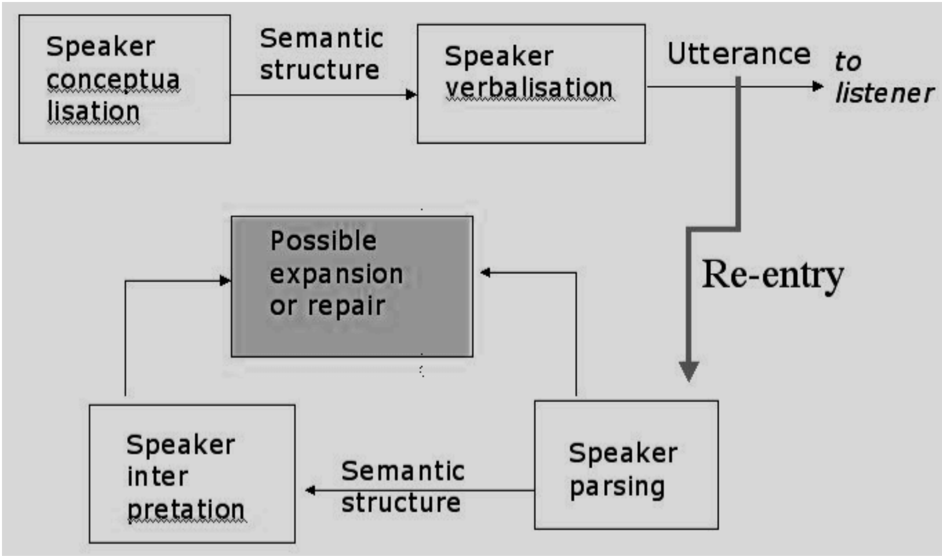

Figure 4.

Output from the language production components is fed back into the language parsing and interpretation components, thus realizing a re-entrant mapping.

production system is very complex and functions automatically and unconsciously. It applies the rules it has and it cannot be planned consciously whether an utterance is going to be adequate from a communicative point of view. The processes of parsing and interpreting the utterance are also automatic and unconscious but they will breakdown or at least detect difficulties if a construction is ambiguous, does not make sense in the present context, or requires more effort than normal. These processes can therefore pinpoint when and where a repair or expansion of the language would be useful and trigger the language invention or language learning subcomponents.

There is plenty of neurophysiological evidence for such a re-entrant system in the human brain. The same left inferior frontal region is activated both for listening to somebody else’s speech and for listening to your own speech — even if this speech is not pronounced (McGuire *et al.*, 1996). Moreover this same region has also been observed to be critical for self-awareness (Stuss, 1991), supporting the argument that inner speech contributes to our sense of self (see below). It is also widely known that people self-monitor and correct their own speech on the fly, which suggests that they must simulate to some extent the hearer’s interpretation of what they are about to say. Comprehension breakdowns or excessive complexity interrupting the automatic processing of language can even be seen by a shift in brain wave patterns (Brown and Hagoort, 1991).

Auditory hallucinations in schizophrenic patients provide further evidence that a re-entrant language system is possible in the human brain but it also shows that this mechanism may fail to function properly or have negative side effects. Schizophrenic patients claim to hear inner voices. They are no longer able to distinguish whether the speech stream produced by their own language production system is internally or externally produced (Frithand Done, 1988). So as with many evolutionary advantages, there are also disadvantages and possible disorders.

2. Emergent case grammars

Let me now elaborate through a concrete example of grammar, why I believe a re-entrant mapping is necessary for bootstrapping a lexical language into a grammatical one. I will focus on case grammar. It is used in natural languages to make event–object relations explicit. For example, in the sentence *The red block pushed the green block off the table*, English speakers use word order, prepositions and hierarchical structure (Subject–Verb–Object) to indicate that the red block is the active element and the green block undergoes the action. Many languages use other ingredients such as case markers (as in Latin or German) or particles (as in Japanese) to express similar sorts of relations. How might such a case grammar arise and how can it be learned? Suppose that speakers only master a pidgin-like protolanguage, in which there are no case markings nor prepositions, and where word order is not a carrier of meaning. They can then only formulate descriptions like: *Push red block green block*.

When speaker and listener are in the same context and observe the same situation, a unique interpretation might still be found by the listener. However when the listener has to interpret the sentence without being in the same context, or when there are two events in the distant past, one where the red block pushes the green block and another one where the green block pushes the red one, the sentence is ambiguous. The role of case grammar consists in disambiguating such situations. But the question is how can such a case grammar arise?

In one of our experiments (Steels, 2002), the agents start with a pidgin-like language without case marking. When a speaking agent encounters a communicative failure on the part of the listener, he runs the sentence through his own parsing and interpretation processes and discovers where exactly the comprehension failure might have originated. For example, if there are two push events, the speaker's parsing and interpretation processes will detect and signal multiple interpretations. The speaker can then fix this by introducing a marking of the event–object relation, and thus a piece of case grammar into his evolving language.

How can the listener learn the meaning of this marking? First, the listener must identify with enough certainty the interpretation that was intended, for example by additional communications. Next he must fill in the missing gaps by hypothesizing that the unknown marker must be related to his comprehension breakdown. By analysing this breakdown the same way the speaker has done, the meaning of the marker can be deduced. Our simulations (which of course contain much more complexity than discussed here) show that a population of agents programmed with these capabilities indeed evolves a shared system of case markings, signalling in abstract terms the event–object relations occurring in their world.

IV: The Society of Mind

Thus far, I have argued that a re-entrant mapping of speech output as speech input is useful — even necessary — for detecting and repairing language communication, and thus for pushing language and its underlying meaning towards greater complexity. Re-entrant mappings are present all over the brain (Edelman, 1989). For example, the areas in the visual cortex receive much more inputs from higher levels than directly from the retina. It is therefore not surprising that re-entry would also be exploited for language. Language communication is so vital and important that it keeps going even if we do not speak.

Once such a re-entrant system is in place, two other things become possible. First of all, it can be used outside the context of communication. It is possible to conceptualize and verbalize thoughts not to communicate with somebody else but to test out ‘how ideas sound’, rehearse future dialogue, submit thoughts to self-criticism, or conceptualize and re-affirm a memory of past experiences. This then could have become the foundation for the creation of a sense of self.

Second, language processing components are able to make use of different bodies of grammatical conventions (and this is already achieved in our computer simulations as well). For example, to speak a foreign language a different set of language rules must be activated, even though the overall architecture of language processing remains the same. We store vast amounts of knowledge about how others speak and how they view their world, which makes us capable of conceptualizing a situation and verbalizing it the way somebody else would. As a consequence, we are able to simulate the voice of somebody else, and thus set up internal dialogs which explore different points of view or prepare us for a conversation with somebody else. This is perhaps the way a ‘society of mind’ as envisioned by Minsky (1985) could have emerged.

At the moment the creative construction and use of representations, both for building a significant model of the self and for preparing interaction with others is completely beyond the current state of A.I. research. In fact, it suggests a whole new avenue of research, which could perhaps build further on the constructionist paradigm which has emerged from Piagetian psychology (Ackermann, 1999) and recent research in cognitive semantics.

V: Conclusions

Normal persons report an ‘inner voice’ which becomes active when one is not speaking overtly. As argued by many authors, this inner voice is an important part of the experience of the self. I have argued that the technical prerequisites for such an inner voice are quite easily established as soon as there is a re-entrant mapping in which output from speech production is internally streamed as input to speech understanding. I believe that this is a necessary prerequisite for the invention and learning of grammatical language, as our computer simulations have shown. Once this re-entrant mapping is in place, it can also be used as a way to ‘listen to oneself’, in other words to have an inner voice through which a self-model can be constructed and tested. Although this self-model is not to be identified with consciousness, it is surely part of the conscious experience and therefore the development of complex language communication may have played a crucial role in the origins of consciousness.

Computer simulations and experiments with grounded robots show that it is technically quite feasible to simulate the effect of an inner voice and that it is in fact necessary to achieve certain mental functions such as the ability to acquire complex grammar. However, just as in the case of attentional and turn-taking behaviour, unravelling the information processing foundations of some aspect of consciousness is only part of the full story. The puzzle remains how the information processing structures which realize certain functions associated with consciousness, can give rise to the first-person experience of consciousness (Chalmers, 1996). There is a difference between an actor who displays a certain emotional feeling like anger or love and a person who really experiences these feelings, even though for an outside observer their behaviour might be indistinguishable. Consciousness is not only a matter of information processing (even though that is certainly part of it) but also a felt first-person experience. I personally believe that robotic models may become more and more sophisticated in capturing some of the information processing aspects of consciousness, but whether this will ever lead to the first-person experience itself is a question I would prefer to leave open at the moment.

Acknowledgement

I am indebted to Owen Holland for stimulating the writing of this paper and for the many comments from reviewers. The thoughts expressed here have been partly shaped by discussions with Susan Greenfield, Lewis Wolpert and Daniel Dennett at the Carlo Erba Symposium on Consciousness held in Milan, December 2001. I am grateful to Professor Colonna and Kathleen White for inviting me to this wonderful meeting. Additional ideas have come through my invited talk at the Conference on Consciousness and Language, organized by the Society for the Scientific Study of Consciousness in Barcelona, June 2002, and from discussions with Israel Rosenfield. I am further indebted to Marleen Wynants for comments that improved this paper.

References

Ackermann, E. (1999), 'Enactive representations in learning: Pretence, models and machines', in *Learning Sites: Social and Technological Resources for Learning*, ed. J. Bliss, R. Saljo and P. Light (Amsterdam: Pergamon).

Austin, J.L. (1975), *How To Do Things With Words* (Cambridge, MA: Harvard University Press).

Breazeal, C. (1998), 'A motivational system for regulating human-robot interaction', *Proceedings of AAAI98*, Madison, WI.

Brown, C. and Hagoort, P. (ed. 1991), *Neurocognition of Language* (Oxford: Oxford University Press).

Chalmers, D.J. (1996), The Conscious Mind: In Search of a Fundamental Theory (Oxford: Oxford University Press).

Clark, H. and Brennan, S. (1991), 'Grounding in communication', in *Perspectives on Socially Shared Cognition*, ed. L. Resnick, J. Levine and S. Teasley (Washington, DC: APA Books).

Crick, F. and C. Koch (1995), 'Are we aware of neural activity in primary visual cortex?' *Nature*, **375**, pp. 121–3.

Dennett, D. (1991), *Consciousness Explained* (London: Penguin Books).

Edelman, G. (1989), *The Remembered Present: A Biological Theory of Consciousness* (New York: Basic Books).

Fodor, J. (1975), *The Language of Thought* (New York: Crowell).

Frith, C.D. and Done, D.J. (1988), 'Towards a neuropsychology of schizophrenia', *Br J Psychiatry*, **153**, pp. 437–43.

Heine, B. (1997), *Cognitive Foundations of Grammar* (New York: Oxford University Press).

Hopper, (1991), 'Emergent grammar', in Traugott and Heine (1991).

Kuroki, Y., Ishida, T., Yamaguchi, J., Fujita, M. and Doi, T. (2001), 'A small biped entertainment robot', in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, Waseda University, Tokyo.

LeDoux, J. (1998), *The Emotional Brain: The Mysterious Underpinning of Emotional Life* (New York: Simon and Schuster).

McGuire, P.K., Silbersweig, D.A., Murray, R.M. David, A.S., Frackowiak, R.S.J. and Frith, C.D. (1996), 'Functional anatomy of inner speech and auditory verbal imagery', *Psychological Medecine*, **26**, pp. 29–38

Minsky, M. (1985), *The Society of Mind* (New York: Simon and Schuster).

Penrose, R. (1990), *The Emperor's New Mind* (Oxford: Oxford University Press).

Rosenfield, I. (1992), *The Strange, Familiar and Forgotten: An Anatomy of Consciousness* (New York: A. Knopf).

Scassellati, B. (1998), 'Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot', in *Computation for Metaphors, Analogy and Agents: Vol. 1562 of Springer Lecture Notes in Artificial Intelligence*, ed. C. Nehaniv (Berlin: Springer-Verlag).

Sperber, D. and Wilson, D. (1986), *Relevance: Communication and Cognition* (Cambridge, MA: Harvard University Press).

Steels, L. (2001a), 'Language games for autonomous robots', *IEEE Intelligent Systems*, September/October 2001, pp. 16–22.

Steels, L. (2001b), 'Social learning and verbal communication with humanoid robots', in *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, Waseda University, Tokyo.

Steels, L. (2002), 'Computer simulations of the origins and evolution of case grammar', Fourth Evolution of Language Conference. Harvard, Cambridge, MA.

Steels, L. and Baillie, J-C. (2002), 'The grounding of dynamic world models by autonomous robots', in *Robotics and Autonomous Systems* (Special Issue on Grounding).

Steels, L. and Kaplan, F. (2001), 'AIBO's first words. The social learning of language and meaning', *Evolution of Communication*, **4** (1), pp.

Steels, L., Kaplan, F.. McIntyre, A. and Van Looveren, J. (2002), 'Crucial factors in the origins of word-meaning', in *The Transition to Language*, ed. A. Wray (Oxford: Oxford University Press).

Stuss, D. (1991), 'Disturbance of self-awareness after frontal system damage', in *Awareness of Deficit After Brain Injury: Clinical and Theoretical Issue*s, ed. G.P. Prigatano and D.L. Schacter (New York: Oxford University Press).

Talmy, L. (2000), *Toward a Cognitive Semantics: Concept Structuring Systems (Language, Speech, and Communication)* (Cambridge, MA: The MIT Press).

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

Traugott, E. and Heine, B. (1991), *Approaches to Grammaticalization. Volume I and II* (Amsterdam: John Benjamins Publishing Company).