

The Problem of Inner Speech and its relation to the Organization of Conscious Experience: a Self-Regulation Model.

Robert Clowes

Centre for Research in Cognitive Science
Department of Informatics
Sussex University
Brighton BN1 9QH
East Sussex
UK
robertc@sussex.ac.uk

Abstract

This paper argues for the importance of inner speech in a proper understanding of the structure of human conscious experience. It reviews one recent attempt to build a model of inner speech based on a grammaticisation (Steels, 2003). The Steels model is compared with a *self-regulation* model here proposed. This latter model is located within the broader literature on consciousness. I argue the role of language in consciousness is not limited to checking the grammatical correctness of prospective utterances, before they are spoken. Rather, it is more broadly activity structuring, regulating and shaping the ongoing structure of human activity in the world. Through linking inner speech to the control of attention, I argue the study of the functional role of inner speech should be a central area of analysis in our attempt to understand the development and qualitative character of human consciousness.

1 Introduction

To introspection, for many of us, our mental life seems to have a constant accompaniment of inner speech. This speech is known in the literature under a number of names such as; the inner voice, the internal monologue, and is sometimes, subsumed into (the more general) stream of consciousness (James, 1890). It may also be linked to the generally pejoratively associated notion of ‘voices in the head’. Understanding the nature of this phenomenon and its functional underpinnings, although of occasional interest in the history of psychology, has, in the last few years drawn the attention of many researchers into mind. There is however, much controversy about the precise nature of inner speech, its epistemic status and possible functional role.

Among psychologists, one means of accounting for inner speech is Baddeley’s articulatory loop (Baddeley & Hitch, 1974),

later rechristened the phonological loop¹ (Baddeley, 1997). This is considered to be a speech related working memory system.

Among philosophers, the notion of inner speech suggests privileged access to mental states, and this, at least in the 20th century, has invited great scepticism. The high-water marks of this scepticism are probably Ryle’s (1949) *The Concept of Mind* and Dennett’s (1991) *Consciousness Explained*. Dennett’s view is complex on this question for although he ultimately doubts the strength of the epistemic warrant that can be given to the narrative stream of consciousness, and especially the subject’s privileged position to report on its contents, he nevertheless argues that the subject’s self-reports should be our starting-point. This is fundamental to his *heterophenomenological* method. This approach advocates

¹ Presumably this re-naming has something to do with thinking of inner speech as primarily an imaged sound, rather than unvoiced speech. The notion of a phonological loop seems to focus on the phenomenology of the passive, rather than active aspect of inner speech.

we need to attempt to offer some explanation of the importance attached to inner speech in phenomenological accounts.

A window into the phenomenology of inner speech is provided by Russell Hurlburt's *Descriptive Experience Sampling* technique (1990). Hurlburt uses an experimental technique in which subjects are cued by a small alarm device at various moments in their day, and then following protocols developed by Hurlburt, write down the details of their mental imagery at the moment that the alarm went off. He argues this technique allows us to systematically sample the qualitative characteristics of reported phenomenology². It also allows us to describe some of the characteristics of inner speech, and inner imagery in general, in a much more elaborated fashion.

The content and form of this reported inner speech seems to be very diverse. Some people report the perception of being the author of voice-like inner speech; others, to hearing voices offering advice or consolation. Sometimes this voice appears to be their own, and sometimes the voice of another person. Some people report merely having the sense of experiencing language-like cognitive episodes without necessarily hearing any voices or having the sense of being the author of this speech. The variety of this speech might serve as some justification for the sceptics, or perhaps just evidence of the complexity and variety of the roles played by speech in our mental lives.

All of these phenomena seem to vary considerably both across individuals, within individuals at different times and places, and with regard to whatever activities they are at that moment engaged in. Hurlburt's

² Although the beeps themselves are random, statistical techniques can be used to understand the distributions of reported mental-events types and indeed correlate them with other types of behavioural measures. (R. Hurlburt & Heavey, 2004)

work reveals much of the contents of consciousness appear to be composed of speech-like episodes. Except in cases of severe psychological disturbance or other abnormal functioning, the inner voice seems to be the constant accompaniment of human conscious life. But can we relate these accounts of the contents of conscious experience to language as vehicle?

Some recent accounts of cognitive role of language have brought to the fore the way that language may play a role, in sculpting, stabilising, and supporting forms of thought which would be otherwise impossible (Carruthers, 2002; Clark, 2004). Trying to forge a link theoretically between the phenomenological and functional aspects of inner-speech has proved so far a difficult task, but it is one upon which some progress has now started to be made.

2 – A re-entrance model of inner speech

Although traditional work on cognitive modelling made much use of more-or-less linguaform internal representations, following (if sometimes implicitly) some version of Fodor's (1975) Language Of Thought hypothesis, it has shied away from explicitly modelling the inner voice (cf. Dennett, 1994). Perhaps this is because of a worry that the inner voice might be either an epiphenomenon or user "illusion" (Dennett, 1991).

Recently work in machine consciousness has begun to treat the phenomenon of inner speech and its possible functional role more directly (Steels, 2003). Steels' earlier work used individual-based models in multi-agent systems to investigate the development of collective lexicons. More recently he has extended these models to attempt to model syntax.

In Steels' newer models agents are able to check the intelligibility of their own sentences by feeding back a prospective utterance through their language interpretation machinery prior to communication. Systems of agent with such *re-entrant* loops appear to be able to self-organise more complex grammars than would otherwise be the case. (Steels, 2003, 2005) Re-entrancy in Steels' models serves the role of checking the intelligibility of an utterance in their own reception systems. Systems of such "self-talking" agents seem to be able to achieve much more stable grammars as a result.

It seems that in order to develop the abilities to use complex syntax, re-entrant loops may be necessary. Steels is thus able to persuasively link re-entrancy to the generation of complex grammars in natural language and perhaps thereby provide a functional role for the inner-voice.

One problem for this work is that the everyday construction of grammatical sentences is usually considered a largely *unconscious* activity. In fact, the construction of grammatically correct sentences is often given as the paradigmatic example of what an unconscious cognitive process is like. Thus, there seems a little *prima facie* implausibility in correlating the phenomenological inner voice with a mechanism whose principle cognitive role is the construction of grammatically correct utterances. While Steels' arguments about the role of re-entrancy in the generation of complex grammars are convincing, arguably however the link with the inner-voice is less well-made.

One important caveat should be put on this observation. Insofar as we are treating the ontogenesis of language in young children, and the problems of developing capabilities to use a language for the first time, it may very well be the case that a large portion of

the child's cognitive resources taken up in assembling and comprehending sentences and possibly they are much more conscious of this. It may turn out that the kinds of activities that Steels models in his experiments might very well turn out to play a central role in the consciousness of young children, and perhaps be the trailblazers for more elaborate forms of conscious inner loops to be developed later in their lives. A further task is to establish links between the Steels model and the account of the inner voice posited by theorists seeking to understand the re-organisation of cognition by language? Arguably his account could be made to fit with some of the recent accounts of language-for-thought that rely on the idea that language allows information to be passed between modules which wouldn't otherwise connect (cf- Carruthers, 2002). As the Steels model seems to have the language production and reception system rather separated from other forms of cognitive activity, it is difficult to say precisely how this relation could be established. Yet if the development of grammar turns out to be linked in this way to a re-entrant cognitive architecture, one can imagine how this architecture could become appropriated by other cognitive functions.

Although the Steels model offers an interesting attempt to show the functional importance of inner speech in order to stabilise the learning of grammars of certain complexity this model may be a special instance of the more general case where self-directed speech serves to scaffold and stabilise a whole range of cognitive functions. Yet could such a system also be linked to the phenomenology of inner-speech and the role of language in consciousness? More work clearly needs to be done in order to establish such a connection.

3 – A self-regulation model of inner-speech

Recent research conducted with Tony Morse (2005)³ demonstrates how an alternative model of self-directed speech, still based on re-entrancy, might relate the inner-voice to a range of broader cognitive activities. The starting assumption for this work is that the cognitive role of language is better understood as one of sculpting or regulating cognitive activity rather than exhaustively representing the world (cf. Clark, 1996). Inner speech could here be seen as serving as a scaffold for developing and sustaining cognitive functions beyond the parsing and construction of meaningful and grammatical utterances.

In our model we compare a series of possible architectures for minimal cognitive agents which have to respond to instructions in order to fulfil externally indicated goals, i.e. moving objects around in a blocks world⁴. Our experiments compare several types of agents with differing architectures, some with word re-entrant loops and some without. All agents are implemented with simple recurrent neural networks that are evolved with a genetic algorithm in order to respond to commands by performing tasks. Some of the agents have architectures that allow the re-triggering of command reception systems internally.

The cognitive architecture of the ‘re-entrant’ agents is arranged such that they can re-use the channels which are being used to signal commands to them to re-

trigger their own behaviours. These channels allow at least the possibility of establishing new control circuits that use the same nodes that have previously been used to receive input from external ‘words’. The thought here is that if there is some advantage to be had by re-using circuits developed to respond to words then the agents will take advantage of this source of useful adaptation. We find this is the case. Even such minimal agents can take advantage of these contingencies to develop word-based modes of self-regulation.

We show that agents with these ‘re-entrant speech’ capabilities (as illustrated in **Figure 1**) perform considerably better on certain tasks. This is explained in greater detail in (Clowes & Morse, 2005). The basic finding is that agents that have architectures allowing the re-use of language for self-regulation achieve higher levels of performance more quickly and can stabilise them for longer than those that do not. Agents that are able to succeed in all task conditions make considerable use of auto-stimulation with words, i.e. they use re-entrant word nodes to self-trigger.

Re-entrance does not function in our models to facilitate merely communicative success or the generation and interpretation of complex linguistic constructions, but in the construction of more viable behaviours. Words here are appropriated in a way that is reminiscent of what Dennett calls auto-stimulation but not as a complex self-question (Dennett, 1991), but as new mode of self-regulation. This work then supplies at least a proof of concept that word-like constructs can be appropriated from a role in regulation from the outside (response to a command) to internal regulation (the agent self-regulating).

But linking such quite basic modes of auto-stimulation with words to inner speech, suggests a rather different picture of its un-

³ A much more detailed examination of this work is now available in my unpublished DPhil thesis.

⁴ NB. This is not exactly a blocks-world in the traditional sense. Rather, agents have extensive sensorimotor couplings with their limited world rather than it being specified in a purely abstract way. The agent architecture itself is an extension of an active vision model reported in experiments by (Floreano, Kato, Marocco, Sauser, & Suzuki, 2003)

derlying nature to that suggested by the Steels model. Inner speech is, I argue, the phenomenological dimension of internalised, word-based self regulation.

The phenomenological appearance of such speech, as speech, depends on it playing a

similar attention focusing role as outer social speech often does. Further, I would conjecture that it relies on the same neural circuits, albeit appropriated for new self-directed functions.

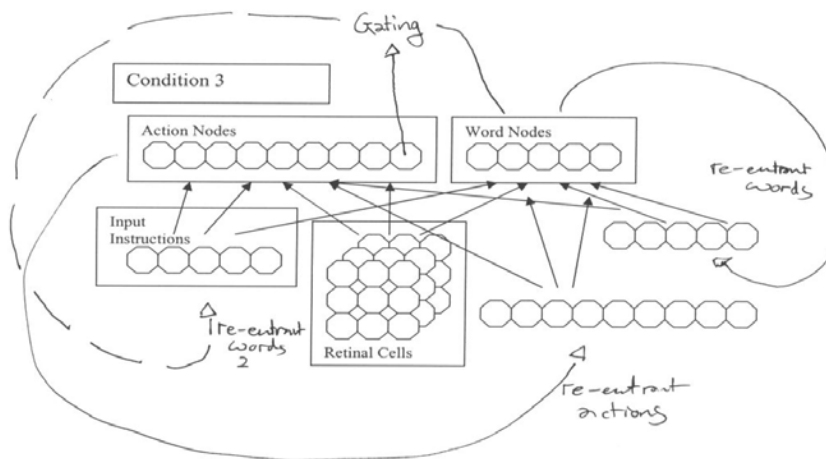


Figure 1 - The diagram shows an outline of the neural-architecture that is used in the experiments. The salient aspect is that when a gating neuron is switched on, activity from the output of the network can be fed back through the nodes that are used as input instructions. More detail on the architecture and some tasks can be found in (Clowes & Morse, 2005). Agents evolved in these conditions develop elaborate self-control loops and develop and stabilise solutions to more tasks than those that do not have such loops.

4 - A functional role for inner-speech

Normal intersubjective speech can certainly play a role in orienting attention, so why not internal speech? A shout in the street can cause an immediate refocusing of attention, e.g., hearing someone shout “mind the car!” as you were about to cross the road, would cause a fundamental reallocation of your attention.

If the inner voice could similarly be linked in some way to the allocation of attentional resources then there is the possibility that it may provide a window into the relationship between higher cognition and consciousness more generally. According to Vygotsky the internalisation of speech forms a

whole new mode of attentional re-organisation.

Vygotsky (1986) emphasized the role of language in the development of control of action and ultimately of attention. His work provides an interesting possible way into the relationship between inner-speech and consciousness by looking at it through a developmental prism.

Vygotsky developed his ideas about the *internalisation* of language in part as a critique of Piaget’s ideas about so-called *egocentric speech*. What Piaget called egocentric speech, and developmentalists tend to call today *private speech*, is a type of speech that children produce between the ages of about 4 and 7. It appears to be addressed toward the self and eventually seems to disappear.

For Piaget this speech occurs toward the end of his pre-operational stage and signifies a still undeveloped ability to take, or imagine, the perspective of others. Social speech was thought to be built from this egoistic basis as children gain more experience that the point of view of others can be different (especially through argument with peers).

A longstanding controversy has arisen amongst developmentalists about the provenance and direction of this speech. Whether it is ultimately a disappearing artefact of early developmental egotism as Piaget argued in his early writings (1926), or alternatively the establishment of the bridge to linguistically controlled higher psychological function (Vygotsky, 1986 - originally 1934), either way this speech does not seem to serve a standard communicative function.

If Vygotsky's theory is correct, then inner-speech has at least its developmental precursors in this particular form of practically oriented speech found in children. If moreover inner-speech once fully internalised could come to play a role in allocating attention then this could provide a strong link between the internalisation of language and the constitution of human consciousness. Understanding inner speech may yet prove to be the royal road to understanding consciousness.

5 – Self-Structuring through internalisation

Much of the theoretical work arguing that language plays a role in consciousness depends on the idea that language reshapes our underlying cognitive mechanisms in some way. Exactly how and to what purpose this functional re-organisation is achieved is currently part of a lively debate.

The potential for re-using language as an addition to the brain's basic modes of organisation is something which is now starting to be taken very seriously in the philosophy of cognitive science (cf. Clark, 2004; Wheeler, 2004).

Dennett (1991) has argued that the development of the self-questioning form of self-directed speech is absolutely pivotal in the construction of human consciousness and its ability to sustain elaborate narrative threads. His view on this seems linked to his position that the form of human consciousness is the effect of installing what he calls a 'serial virtual machine' on parallel processing hardware. A range of accounts of the functional role of inner speech and its relationship with consciousness have also been put forward (Carruthers, 2002; Clark, 1998; Frawley, 1997) which seek to expand upon or restructure in various ways the sort of picture developed by Dennett. Although it seems possible that episodes of inner speech are epiphenomenal and fulfil no functional role in the organisation of consciousness, it is certainly too early to rule out the contrary possibility.

One can derive a further link between self-directed speech and the functional structure of consciousness from the psychopathological literature. Evidence seems compelling that the collapse of a normal inner voice in disorders such as 'thought insertion' is often correlated with catastrophic breakdowns for the organisation of individual consciousness (R. T. Hurlburt, 1993; Stephens & Graham, 2000). Disorders such as schizophrenia are sometimes theorised as control disorders and this idea gives us a way into establishing a possible link with the functional role of internalised speech (Gallagher, 2000). It points towards some quite central role for self-directed speech in the organisation of human consciousness, if not necessarily along the lines of Dennett's model.

One difficulty with this idea is that it is still very unclear at the level of sub-personal cognitive architecture how language can come to play the types of roles that are being ascribed it by the consciousness theorists. Yet there is a dearth of cognitive models that even attempt to show how such a reorganisation might happen⁵. However, it is possible to further analyse the model described above to give some insight into how attentional control through language internalisation might be established.

The model presented here gives one suggestion as to how the sorts of complex modes of self-regulation that seem bound up with human consciousness can get underway.

The simulation work with minimal cognitive agents shows that the re-use of public symbols in re-organising the ongoing activities of self can have cognitive benefits. These appear to go beyond being able to interpret and sustain more complex languages. Rather the internalisation of language in these models has more to do with the restructuring ongoing situated action.

Analysing the models further we found that the development of the ability to re-use a system of commands appears to move through broadly three control regimes.

1. Agents develop the capacity to respond to instructions. At this stage of development agents might be described as passive and do not use self-directed instructions very much.

⁵ Despite these lacuna in more general work on cognitive modelling and the role of language some interesting work linking linguistic and cognitive function is starting to be done (Sugita & Tani, 2002). This work however encompasses quite a distinct formulation of the idea of a role for language in cognition as does the work reported here.

2. Agents start to auto-stimulate with instruction nodes. This regime of self-control tends to produce ineffective and unstable systems of activity, (e.g. agents can sometimes perform the tasks well but very often do not).
3. Finally agents develop much more robust forms of self-control that rely on the ability to use new regimes of action made available by the self-directed loops.

Can these results be linked with Vygotsky's ideas about the establishment of new regimes of self-control through the internalisation of speech?

Vygotsky – to some extent developing the ideas of the Gestaltists⁶ - argued that the development of self-directed speech was an form of self-prompting by which children come to de-centre and move themselves from one domain of situated activity (or as he might have termed it practical thought) to another. He saw this development as being centrally involved in the establishment of self-control and attention-regulation that are characteristic of human consciousness.

The work discussed above gives us a possible way of understanding the neural-dynamics underlying the establishment of this linguistic self-regulation.

6 – Inner speech and the modelling of consciousness

Notwithstanding current attempts to develop work in synthetic phenomenology (Chrisley & Holland, 1994), for now⁷, hu-

⁶ Gestalt psychologists wrote a great deal on the problem of insight and how it was that a problem might suddenly be restructured such that it appears in an entirely new way. Kohler was one that held that tools could play a role

⁷ Perhaps forever, cf, (Nagel, 1974)

man consciousness is the only type of consciousness which we know intimately. It seems unlikely that we can afford to ignore the relevance of the role of language in attempts to model it in machines, not to mention the project of building actually conscious machines.

Theorists as diverse and as historically distant as Vygotsky and Dennett have argued that self-directed speech plays a central role in the organisation and even the construction of human conscious experience. Work by Hurlburt and others appears to show that conscious experience abounds with episodes of internal speech.

If they are right and we are serious in our attempts to understand human consciousness with synthetic techniques, then we need to develop more advanced and explicit models of the role language might play in its functional organisation. The hypothesis defended here about the functional role of internalised speech is that it is a tool for the focusing or re-focusing of attentional resources.

Inner speech then appears to be of central importance because it gives an agent the capacity to restructure not just the external world but also itself. External activity in this way becomes redeployed toward inner restructuring. Simulation models such as those discussed above give us a unique mode of developing an understanding of the functional changes that underlie such a transition.

This internalisation model of self-directed speech can be used to provide an explanation of how language plays a role in creating the regimes of complex self-control and attention-regulation that are central to the sorts of consciousness that humans have (cf Donald, 2001). It does not attempt to address the question of why any experiences are conscious at all. However, it may allow

us a new vantage point on their qualitative character.

According to the sensorimotor approach or ‘skill theory’ of conscious experience, “experience is not something we feel but something we do” (O'Regan, 2001). The character of perceptual experience, according to this theory, is given in the mastery of sensorimotor contingencies. These contingencies of self have their own governing laws just as any other complex physical system. Developing a mastery of these laws through autostimulation-with-words might be considered akin to the development of a new perceptual modality.

This mastery of the mechanisms of autostimulation-with-words affords the refocusing of one's own attention on self. This exercise of the contingencies of self can therefore be linked, more generally, to the qualitative analysis of consciousness in terms of sensorimotor contingencies (cf O'Regan & Noë, 2001). Understanding this refocusing of attention might help us explain the uniquely human mode of the self's perceptual presence.

References

- Baddeley, A. (1997). *Human Memory Theory and Practice*. Hove, UK: Psychology Press.
- Baddeley, A., & Hitch, G. (1974). Working Memory. In G. A. Bower (Ed.), *Recent advances in the psychology of learning and motivation*. New York: Academic Press.
- Carruthers, P. (2002). The Cognitive Function of Language. *Behavioral and Brain Sciences*, 25(6).
- Chrisley, R., & Holland, A. (1994). *Connectionist synthetic epistemology: Requirements for the development of objectivity* (No. 353): COGS CSRP 353.

- Clark, A. (1996). Linguistic Anchors in the Sea of Thought? *Pragmatics And Cognition*, 4(1), 93-103.
- Clark, A. (1998). Magic Words: How Language Augments Human Computation. In P. Carruthers & J. Boucher (Eds.), *Language and Thought. Interdisciplinary Themes* (pp. 162 - 183). Oxford: Oxford University Press.
- Clark, A. (2004). Is language special? Some remarks on control, coding, and co-ordination. *Language Sciences*, 26(6), 717-726.
- Clowes, R. W., & Morse, A. (2005). Scaffolding Cognition with Words. In L. Berthouze, F. Kaplan, H. Kozima, Y. Yano, J. Konczak, G. Metta, J. Nadel, G. Sandini, G. Stojanov & C. Balkenius (Eds.), *Proceedings of the 5th International Workshop on Epigenetic Robotics*. Nara, Japan: Lund University Cognitive Studies, 123. Lund: LUCS.
- Dennett, D. C. (1991). *Consciousness Explained*: Penguin Books.
- Dennett, D. C. (1994). The Role of Language in Intelligence. In D. C. Dennett (Ed.), *What is Intelligence*. Cambridge: Cambridge University Press.
- Donald, M. (2001). *A Mind So Rare: The Evolution of Human Consciousness*. New York / London: W. W. Norton & Company.
- Floreano, D., Kato, T., Marocco, D., Sauser, E., & Suzuki, M. (2003). *Active Vision & Feature Selection: Co-development of active vision control and receptive field formation. Complex visual performance with simple neural structures*. Retrieved 30 June 2004
- Fodor, J. (1975). *The Language of Thought*. New York: MIT Press.
- Frawley, W. (1997). *Vygotsky and Cognitive Science: Language and the Unification of the Social and Computational Mind*. Cambridge: Harvard University.
- Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*.
- Hurlburt, R., & Heavey, C. L. (2004). To Beep or Not To Beep: Obtaining Accurate Reports About Awareness. *Journal of Consciousness Studies*, 11(7), 113-128.
- Hurlburt, R. T. (1990). *Sampling Normal and Schizophrenic Inner Experience*. New York: Plenum Press.
- Hurlburt, R. T. (1993). *Sampling inner experience with disturbed affect*: Plenum Press.
- James, W. (1890). *The Principles of Psychology*.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83, 435-450.
- O'Regan, J. K. (2001). Experience in not something we feel but something we do: a principled way of explaining sensory phenomenology, with Change Blindness and other empirical consequences.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24.
- Piaget, J. (1926). *The Language and Thought of the Child*: Routledge and Kegan Paul.
- Ryle, G. (1949). *The Concept of Mind*. Chicago: The University of Chicago Press.
- Steels, L. (2003). Language Re-Entrance and the 'Inner Voice'. In O. Holland (Ed.), *Machine Consciousness*. Exeter: Imprint.
- Steels, L. (2005). Constructivist Development of Grounded Construction Grammars.
- Stephens, G. L., & Graham, G. (2000). *When Self-Consciousness Breaks*: MIT Press.

- Sugita, Y., & Tani, J. (2002). A connectionist model which unifies the behavioral and the linguistic processes. In M. I. Stamenov & V. Gallese (Eds.), *Mirror Neurons and the Evolution of the Brain* (Vol. 42).
- Vygotsky, L. S. (1986). *Thought and Language* (Seventh Printing ed.): MIT Press.
- Wheeler, M. (2004). Is language the ultimate artefact? *Language Sciences*, 26(6), 688-710.