AI & Soc (1995) 9:116–137

AI & SOCIETY

Simulating Conversations: The Communion Game

Stephen J. Cowley[1] and Karl MacDorman[2]

[1]*Department of Linguistics, and* [2] *Computer Laboratory, Cambridge University, Cambridge, UK*

Abstract: In their enthusiasm for programming, computational linguists have tended to lose sight of what humans *do*. They have conceived of conversations as independent of sound and the bodies that produce it. Thus, implicit in their simulations is the assumption that the text is the essence of talk. In fact, unlike electronic mail, conversations are acoustic events. During everyday talk, human understanding depends both on the words spoken and on fine interpersonal vocal coordination. When utterances are analysed into sequences of word-based forms, however, these prosodic aspects of language disappear. Therefore, to investigate the possibility that machines might talk, we propose a *communion game* that includes this interpersonal patterning. Humans and machines would talk together and, based on recordings of them, a panel would appraise the relevant merit of each machine's simulation by how true to life it sounded. Unlike Turing's imitation game, the communion game overtly focuses attention, not on intelligence, but on language. It is designed to facilitate the development of social groups of adaptive robots that exploit complex acoustic signals in real time. We consider how the development of such machines might be approached.

Keywords: Adaptation; AI; Conversation; Coordination; Imitation game; Language; Machine learning; Turing test; Simulation

Introduction

Early trailblazers in AI viewed the prospect of axiomatising thought with heady optimism. It was supposed that insight could be gained into human understanding by simulating the performance of intellectually demanding tasks. Despite the success of programs that do such things as play chess or prove theorems, these insights have failed to materialise. Academics working inside the field have increasingly joined those from other disciplines in challenging the foundations of symbolic AI (Winograd and Flores, 1986; Dreyfus, 1972, 1992). There have even been calls to reconceptualise AI in terms of very simple neurone-like units (Smolensky, 1988). Meanwhile, AI researchers have grasped the difficulty of designing machines that get about in the world by themselves. This difficulty was at first unexpected because people and their animal predecessors do so with ease. In response to these findings, some have argued that instead of starting with, for example, language and problem solving,

research should begin with the simulation of what simple creatures like insects do (Brooks, 1991a,b). Others have noted that, because an engineer's understanding is limited, it might be more fruitful to evolve intelligent behaviour than to design it (Sims, 1994).

In that we are concerned with what people actually do, this paper is part of what has recently been called the second cognitive revolution (Harré and Gillet, 1994). Further, by adhering to the revolutionary party, we are challenging one of the remaining bastions of the symbolists: language, we argue, should be modelled as *activity*. Our argument boils down to the familiar one: it is often useful to simulate what is biologically simpler before what is evolutionarily complex. Accordingly, our heuristic involves proposing an expectational model of nonsymbolic aspects of language. Our position, however, is also revisionist: we believe that language and especially written, academic language, also depends on manipulating symbolic forms. Our allegiance to the revolutionary party, then, is traceable only to our conviction that public, nonsymbolic aspects of linguistic activity need to be understood. We reject Bolshevik assertions that the price of developing machines that behave intelligently is the abandonment of all mediating representations.

In making our case, we focus recent insights about conversation around discussion of a single exchange. This provides the basis for arguing that, if machines are to communicate in a human-like way, what is primary in terms of evolution – the prosodic and interpersonal aspects of these events – must not be ignored. To engage in conversational activity, above all, machines must be enabled to adapt to individuals in specific circumstances (Bateson, 1988). Accordingly, to evaluate progress in machine simulation of language, we propose that Turing's imitation game be updated as a *communion game*. Finally, suggestions are put forward concerning how to set about developing adaptive robots able to play it.

Conversations, the Unsaid and Individual Differences

When language is pictured as a system of elements and rules known to speakers, it is imagined that communication depends primarily on the internal structure of what is said. To engage in talk, it is supposed, requires above all the ability to combine linguistic forms in accordance with the same grammatical and logical processes that lend themselves to interpreting speech. Language, in short, comes to be envisaged as an autonomous system that can be described independently both of persons and of how words are spoken. Although this view is counter-intuitive, it has influential defenders including Chomsky (1965, 1986, 1988) and Fodor (1975).

A primary achievement of research in the field of pragmatics has been to show that regardless of whether the mind-brain contains an autonomous language module, language activity cannot be adequately described (let alone explained!) unless language users are brought into account (see Mey, 1993). This is because utterances are not understood person-neutrally. On the contrary, they are construed by particular individuals in specific circumstances and in relation not only to what is said but also to much that remains unsaid. While lexico-grammatical properties undoubtedly contribute a great deal in the course of communication, conversations also depend on what has been termed *interactive intelligence* (Levinson, in press).

The reason for this is readily grasped. While actions can be modelled in terms of plans, goals, and intentions, their interpretation cannot. There can be no determinate way of *inferring* premises from conclusions (Levinson, in press). Even in logic a conclusion *P*, tells us nothing about underlying premises. This is equally true for conversational events. Consider the following utterances:

B: *I have put the glue*

It is unimaginable that anyone could infer premises leading to the issuance of this utterance. It is equally unimaginable that anyone could predict premises that would come to mind as this utterance is heard by B's interlocuters. Certainly, no digital computer relying on speech processing models could do this. Nor indeed would the device be much helped if, by some miracle, it came to "know" that the individuals involved were father, son and daughter, and they were together making model aeroplanes. This is because, contrary to what a reader may be tempted to think, the main point of B's utterance is not to inform anyone that she has put the glue on a particular part of the plane.

Even if a transcription of the entire conversation were supplied, such a textual context does not provide adequate grounds for drawing many inferences about what is going on. In fact, B's utterance was immediately preceded by one spoken by her brother:

A: *Daddy you're supposed to help me with my plane*

Since B's utterance is an example of what a logician would deem a nonsequitur, this may seem surprising. Certainly, from a speech processing point of view, this makes the events particularly difficult to understand. Furthermore, it shows that knowledge context, if limited to the immediate verbal setting and the basic background information, is insufficient to take us very far in understanding conversations.

By contrast, from the perspective of an English speaking observer the discontinuity of this fragment makes it perfectly *easy* to understand. Like the participants themselves, human observers base their understanding on insights drawing, not just on words spoken and background information, but also on *what exactly they hear*. As soon as a human can consider phonic patterning, the exchange in question loses much of its mystery. A speaker's understanding of events is heavily dependent on prosodic, temporal, metrical, and voice quality patterning.[1] This is hardly surprising: were this information not crucial to humans, how could babies, who know no words, ever set off on the path toward spoken language? (see Locke, 1993). To the extent that infant vocalisations express understanding, they must be *entirely* dependent on aspects of speech that are not word-based.

How then, does the prosodic information embodied in the utterances contribute to understanding of what is happening? To grasp this, it is necessary to consider what strikes an attentive listener. One particularly salient fact is that B's utterance is timed to coincide with the last centiseconds of the major pitch accent occurring on *my*. This temporal dimension can be shown as follows (where ↑ signals onset of overlap):

A: *Daddy you're supposed to help me with my plane*

B: ↑ *I have put the glue*

Similarly, the observer would be struck by the lovey-dovey tone of the first utterance, characterised by high pitch and bouncy rhythm. These effects depend, in the main, on falsetto voice, lip rounding, and a babyish lisp occurring during slow speech. Indeed, the auditory effects are so striking that even a person entirely unfamiliar with English could hardly fail to notice how much they contrast with the serious tone of the sister's speech. Anyone familiar with how people typically behave in families would be able to grasp that *B* was timing her speech *to disrupt her brother's interaction with their father.*

To recognise that *B* is disrupting what *A* is doing provides a first inkling of what is happening. In gaining fuller appreciation it is useful to employ acoustic analysis. An observer looking at, say, the physical dimension represented by fundamental frequency (F_0) would find that B's speech was not only exquisitely timed. In fact, *B* pitches her voice so that, at the moment of overlap (the point marked by ↑) she causes maximum disruption: as she interrupts her brother his pitch is falling through a level of 355 Hz and, at this precise moment, her interruption rises from an audibly very similar level at 345 Hz. This interpersonal pitch matching is most unlikely to be a coincidence. To cite but one piece of evidence, although the boy's falsetto (*A*) involves an unusually high fundamental frequency, his sister matches him in this respect (for discussion of similar cases, see Cowley, 1993, 1994). These facts, together with others – unsurprisingly the sister speaks loudly – can be readily interpreted. Similar psychological acumen is required in recognising that the brother and sister are competing for their father's attention. In so doing, the sound of their voices plays a significant part. To understand the utterances, one needs to recognise that the interruption depends on picking up on what another person is saying, matching vocal harmonies, and speaking so as to blot the other person's words.

Communicative effects achieved by extremely fine coordination have been documented through the prosodic domain.[2] Voices set up audible parallels and contrasts over rhythmically and melodically defined units, words, syllables and, as here, single instants of speech. Voices, moreover, parallel and contrast each other simultaneously, exploiting a range of qualitative, rhythmical, and melodic characteristics (Cowley, 1993, 1994). Patterns of coordination are set up in, for example, voice quality, loudness, rapidity and pitch. Nonsegmented aspects of speech are crucial to understanding what people mean with the words they say. In humans, as in other vertebrates, vocal communication depends not just on what is thought but also on how individuals coordinate their bodies (Cowley, in press).

Starting with an Artificial Model of Communication

It would be difficult to construct an argument for ignoring phonetic information like that noted. Nevertheless, that is what typically happens when prosodic aspects of speech are described.[3] Within linguistics, as well as in computational applications of prosodic models, it is almost invariably assumed that what matters are forms that characterise aspects of what people *normally* do. Obviously enough, this is justifiable in describing neutral spoken prose.[4] It is equally legitimate in designing machines for producing speech-like output or transcribing words spoken aloud.[5]

Interactive intelligence, by contrast, typically depends on expectations that are set up and violated in real time. What matters is often less a question of what is

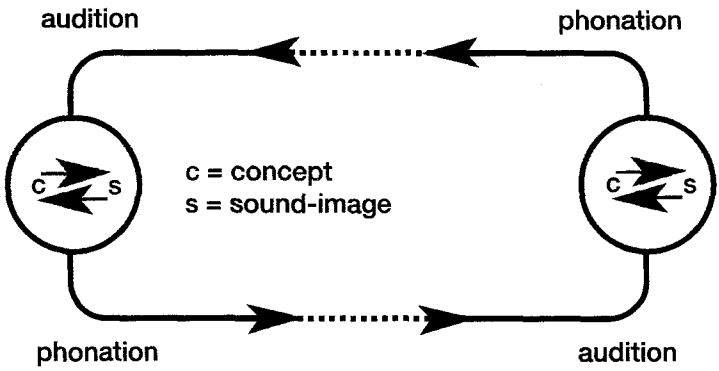

Fig. 1. Saussure's speech circuit (1959, p. 12). According to Saussure, phonation is associatioed with "the representations of the linguistic sounds" (p. 11). In modern terminology this corresponds to a linear sequence of phonological units defined for a language-system. The model assumes that, from participants and observers' perspectives alike, phonation and audition are of identical value and that prosodic (e.g., pitch, rhythm) and voice-quality information contribute nothing to the contact made.

normally done than of what happens between people on particular occasions. Interpersonal aspects of meaning hit people as they happen and can only comparatively rarely be traced to what, in a strict sense, is being affirmed. They depend on striking a chord in the listener. Just as the metaphor suggests, prosody plays a major role in this. Self-evident as this may be, the matter is typically passed over in silence. There is no obvious reason for this state of affairs: no simple justification can be provided for paying less attention to what people mean than to what they say. This bias derives from neglect of language users, and pragmatics has flourished because of its attempt to counter it.

We adopt the view that linguists and computer scientists alike have tended to focus on what is said because they accept assumptions embodied in *speech circuit* models of communication (see Fig. 1).[6] These models can be traced back at least as far as Saussure (1916, 1959, see especially p. 11). Their popularity spread especially with Ogden and Richards' *Meaning of Meaning* (1923), Bloomfield's *Language* (1933), and Shannon and Weaver's theory of "information" (1949).[7] Applied to conversation, speech circuit models wrongly assumed that people took turns at speaking so that shared knowledge of an underlying language-system could provide a common understanding of linguistic *forms*. From an observer's perspective at least, communication was considered reducible to the coding and decoding of signals.[8]

As Harris in particular has argued, this model distorts our understanding of language-systems (1981, 1987). Independently, Krebs and Dawkins have taken the view that related formal models fail to provide insights into animal communication (1984). It is, therefore, unfortunate that computational linguists have so often allowed assumptions implicit in the speech circuit to influence their work.

The problem with the speech circuit model is that it oversimplifies: it focuses not on what *happens* during conversation, but on what might underlie it. This is not to deny, of course, that this emphasis raises extremely important questions. However, what our simple example has shown is that there are *other* equally important questions.

One can also ask what people *do* as they engage in conversation. The problem with the notion of input as text implied by speech circuit models is that it is too crude for describing behaviour of complex agents. B's speech is structured both with respect to what is said *and* to fit the phonetic and temporal characteristics of her brother's coinciding utterance. In their exchange, prosodic factors are more important than the words spoken. How B's utterance intermeshes with her brother's is crucial to what she means. To adopt the speech circuit model is to lose sight of this.

One major problem with the speech circuit, then, is that it dismisses vocalisations as nonlinguistic. Language is distinguishable from behaviour: our physical capacities are separated from our intellectual ones. This perspective emerges especially clearly in Chomsky's work, where grammars are hypothesised to be autonomous mental systems. While this picture enables communication to be imagined as deriving from the processing of rules and representations, in pragmatic terms it is deeply unsatisfying. By separating language from its users, the crucial communicative role of the unsaid is obscured.

It should not be thought that this unsatisfactory picture can simply be traced to the cognitive turn of the 1960s. While it fits the view that a language is a quasi-computational system, its origins go back further. Indeed, as noted, it was the founder of structural linguistics who established both the speech circuit and the accompanying doctrine of linguistic autonomy. Saussure (1916) is completely open about doing this. His motivation was to establish linguistics, in his words, as a science. This move enable him to present languages as essentially unchanging and atemporal "objects". Many linguists have followed Saussure in this. However, the fact remains that no abstract object can be squared against the sorts of observations outlined above.[9] Any such reification is bound to raise what Carling and Moore (1982) refer to as the problem of congruence: none can be correlated with the sorts of *interpersonal* events that are enacted during ordinary conversations.

Linguists who accept the speech circuit model rarely consider language in any basic pre-theoretical sense. They seek to describe, not utterances, but underlying abstractions. In post-Bloomfieldian terms, behaviourists and cognitivists alike conceive of language in relation to configurations of *linguistic forms*. Since they are only forms, they can only be associated with abstract semantic representations. Although these properties, what Lyons calls the descriptive meaning of expression, typically have some bearing on events, only rarely can they explain them. Thus, in the example, while B indeed informs her father that she has put glue on the part, she is primarily concerned with disrupting her brother's behaviour and winning her father's attention. This has next to nothing to do with the words spoken: the event depends less on the message than on the accompanying metamessage (Bateson, 1979). In speaking we are acting; it is we, not our propositions, that refer (Lyons, 1977; Strawson, 1950). Likewise, people, not linguistic forms underlying their utterances, do the complex things that language enables them to do.

This might be unimportant were it not that, from evolutionary and developmental points of view, conversations form the basis of all human language. Especially where these take place in face to face encounters between members of the same community, conversations represent the most basic forms of language, or in Lyon's terms, *ordinary language behaviour*. From this behaviour, as he notes, "All other uses and manifestations of language, both written and spoken, are derived in one

way or another (Lyons, 1977, p. 64). Were this not true, it might be less important that understanding depends not just on the words spoken but also on the individuals concerned and on everything that remains unsaid, and it would be legitimate to develop simulations of language on the model of, say, electronic mail. The fact of the matter, however, is that human languages – and how people mean – are not learnt at computer terminals or in books: they are learnt while participating in a flow of interpersonal prosodically-mediated behaviour.

It is, of course, because the speech circuit model can be applied to derived forms of language, such as electronic mail, that it retains its apparent plausibility (especially for nonlinguists). Indeed, it is not difficult to see why it has been universally adopted in language and speech processing. What is remarkable, however, is that this same model seems to have been applied in the majority of computer applications purporting to come to terms with action and behaviour. This is true not just of symbolic AI but also of many nonsymbolic uses of computing technology. Just as linguists have tended to focus on hypotheses about what may *underlie* acts of speaking, computer scientists have focused on what could underlie, say, problem solving or various kinds of learning. An over-emphasis on behaviour and especially on knowledge, we believe, blocks progress in developing many kinds of machines. Above all, it leads to the passing over of interesting questions concerning how information spreads between complex environments and artificial systems. This overemphasis we find echoed in Turing's belief that it would be useful to separate the intellectual and physical capacities of man (1950, p. 434). Inadvertently, his imitation game was designed in a way that promulgated the speech circuit model of communication.

Revising the Imitation Game

In 1950 Turing wrote, "Conjectures are of great importance, since they suggest useful lines of research" (p. 442). Few conjectures have been more influential in computer science than those underlying the so-called Turing Test. In fact, what Turing proposed was an *imitation game* in which a man competed with a woman at simulating a woman's words in conversation.[10] He predicted that by the year 2000, a computer could be programmed to play the man's part so well that, in a five minute conversation, an average interrogator's chance of identifying the woman would not exceed 70%.

Since its inception, the imitation game has exerted a major influence on artificial intelligence. Nobody would doubt that it has contributed much to the progress and popularisation of computers. This is witnessed, for example, by the many attempts made to simulate written language communication in limited domains (e.g. Winograd, 1972; Schank and Abelson, 1977). There is even a yearly competition based on Turing's game (Shieber, 1994). At the same time, doubt has been cast on the view that the game provides a litmus test of intelligence and indeed on whether it makes sense to define intelligence operationally (Gunderson, 1971; Block, 1978, 1981; Searle, 1980; Karelis, 1986; Bieri, 1988; Hauser, 1993). In spite of these arguments, the imitation game is sure to continue to have a major impact on research.

Unlike many who have discussed the Turing Test in philosophical terms, we do not take the view that the game should be abandoned because it does not tell us

whether machines could think. Rather, our grounds for focusing on the game are linguistic: we do not believe that much further progress will be made in simulating human communication unless the game is updated to bring biologically simpler aspects of language into account. More specifically, we believe that much can be learned from simulating the kinds of behavioural intermeshing that are ubiquitous during informal conversations. Simply, we take the view that the test retains too much influence over the progress of intelligent machines to be left in a form dictated by 1940s technology. Furthermore, by proposing a new game in which the simple textual conception of input is abolished, we believe we are writing in the spirit of Turing's original work. In freeing researchers from the constraints of the speech circuit, our principle objective is to encourage them to explore new areas of inquiry so as to stimulate the development of computing machinery.

Following Turing we adopt the view that informal tests of machine ability provide more challenging goals for computer science than those set by tests deriving from philosophical or psychological theories. In short, we agree that simulation (or, in his words, *imitation*), provides a valuable means of establishing "facts" that support conjectures (p. 442). This is why we too frame the issues in empirical terms.

Facts and Machines: An Artificial Model of Communication

From a pragmatic point of view, the main weakness of models based on the imitation game is that they reduce language to sequences of symbols and, in so doing, disregard both individuals and all that remains unsaid. By assuming that behaviour depends on such a simple kind of input, they omit any reference to aspects of a background of experience that are not amenable to codification. These models, we suggest, fudge our grasp of ordinary communication.

It is not difficult to see why Turing designed his imitation game so that only sequences of symbols had any part in events. His reasons were no doubt largely historical. In the 1940s, technology was such that computer operators had to act within narrow constraints. Further, given the intellectual climate of the times, it was natural enough to assume that little would be lost by treating spoken words as tokens associated with linguistic forms. Thus, in the imitation game, an interrogator interacted with subjects by typing on a teleprinter[11] and reading its output.

It is a tribute to the success of digital computers that the artificiality of this input-output model no longer strikes many readers. Systems developed to fit these devices generally interact with humans only to the extent that *we* actively construe the output provided by the machine. In the imitation game, therefore, the needs of the device compel participants to use the computer-friendly teleprinter. Thus, the activity of human and machine alike is presented as giving rise to what linguists[12] usually conceive of as *meaningful* sequences of symbols (word-based forms). From the computer's point of view, by contrast, the symbols current AI systems process are not, in any ordinary sense, meaningful. They are merely symbol strings (arbitrary sequences of linguistic forms). The ambiguity of *symbol* leads to untold confusion.[13] In practical terms it means, above all, that designing computers to engage in linguistic activity has been conceived almost entirely in terms of developing programs for interpreting symbol strings.

As argued above, there are important reasons for rejecting this model. During conversations people react to what is said while shaping their actions with respect to what other individuals are doing. They react to different persons differently and, by so doing, confirm that the unsaid is as important in communication as are the words spoken. There is, moreover, abundant evidence that these factors are crucial to language development (e.g. Halliday, 1978, 1979; Trevarthen, 1979, 1986; Locke, 1993). To come to terms with them, therefore, what are needed are machines that do not work through symbolically mediated interfaces. Such devices, moreover, must adapt to different individuals and, on the basis of these encounters, develop communicative expectations of their own. Rather than assume that subjects and interrogators are like artificial devices, engineers must be encouraged to develop machines, that, like people, deal with nonsymbolic aspects of language.

If this is to be done in a way of which Turing would have approved, it is necessary to revise his imitation game. The consequence of abandoning the assumption that humans characteristically undertake "symbolic interpretation" is that we need to develop adaptive robots.[14] To evolve their own systems of symbols, they will need to exist in social groups and will, obviously enough, have to do much besides symbol manipulating. In short, in Harnad's (1990) terms, what is needed is a way of getting to grips with how symbols are *grounded* (see Sommerhoff and MacDorman, 1994, section 5.2). Below, we propose that this can be achieved if Turing's imitation game is replaced by a version more suited to the concerns of current AI research.

Proposing a New Game

Turing's imitation game encourages researchers to design machines that *simulate* conversation. A machine's conversation simulates a human's to the extent that their activity is observably similar.[15] Thus, although there can be no simple, principled answer to the question of what is to count as a simulation, people can readily judge not only whether *X* simulates *Y*, but also whether *X1* simulates *Y* better than *X2*. To make such a judgement is to say that one simulation rings more true than another. Indeed, a major virtue of Turing's test is that it provides a means of coming to grips with this question.

In principle, there can be no doubt that talking machines will simulate everyday conversations better than ones outputting standardised word-based forms. While interpreting and producing such forms is important in conversing, we have seen that this is the epilogue to a more intricate biological story. The imitation game will be enhanced when it can be revised to promote the design of machines able to exploit, in a human-like way, continuous fluctuations in phonetic signals. Bearing in mind observations akin to those described above, we put forward two main proposals concerning how the game should be revised.

The first of these is that we drop Turing's requirements that only digital computers be considered as suitable players in the game (p. 436). As is evident enough, especially if these machines come into contact with the world only through a keyboard, they are ill-suited to dealing with biologically simple aspects of language. In proposing our updated version of the game, we expect successful players to draw on different kinds of technology – the robotics required for complex motor activity, parallel processing methods that lend themselves to pattern recognition, and symbol

manipulating techniques for dealing with many word-based aspects of languages. Indeed, unlike Turing, we see no need to stipulate in advance what kind of machines should take part in the game.

Our main proposal, however, is that a *communion game*[16] be substituted for the original imitation game. Instead of placing a machine at odds with a woman who, along with an interrogator, is trying to expose it as an impostor, machines will be allowed to talk with different people informally. These conversations will be recorded and submitted with entirely human conversations to a panel of ordinary people to judge *qualitatively* how true to life they sound. The results from testing simulations in this manner could be used as a methodological tool for their refinement.

Our updated test is not intended to deliver a verdict on whether machines can think. Indeed, far from being a litmus test, it is seen as giving new impetus to Turing's conjecture that machines engage in conversation. The communion game is designed to establish how closely this activity can be brought to resemble ours. Were it possible to build talking machines, it goes almost without saying that an unprecedented number of applications awaits them. Our hope is that, by emphasising that the thinking manifest in talking has an interpersonal and somatic component, our test will stimulate the development of adaptive robots that provide insight into interpersonal activity.

A Defence of the Communion Game

Our communion game couples the goal of getting computers to communicate symbolically with that of developing machines whose vocalisations achieve interpersonal aspects of meaning.[17] Machines taking part in it will handle analogue properties of utterances by setting up and responding to finely coordinated vocal patterning. The goal set by our game is quite unlike that of developing machines able to convince an interrogator that they are human, perhaps by having programmers key in "knowledge".[18] We are proposing the development of machines that *talk*. The key design criterion is whether these can match our vocal adaptability. If they are to relate to people as humans do, they have to consider individual differences. In setting this goal, one crucial to evolution, it will be useful to begin by creating machines that adapt uniquely to each other in functional ways.

The communion game avoids some of the red herrings that have haunted the imitation game. Thus, for example, we have done away with the interrogator. Intense questioning is inappropriate for determining whether a machine can converse like a person. Many humans would not cope with inquisitorial dialogue, and we see no point in encouraging designers to program machines to "pretend" to be human.[19] While the question remains an empirical one, we would expect machines taking part in the game to display interpersonal dynamics similar to those observed in vertebrates. Although their design would be functional, from an outsider's point of view, we would expect their behaviour to appear expressive.

For similar reasons, we have remedied the problem of cultural bias arising in Turing's test (French, 1990, p. 54). The communion game neither requires that human and machine players speak the same language, nor that the machine be familiar with human customs. What matters is that it relate to participants in a human-like way. However, just as we would expect a nondysfunctional person to adapt to other individuals (Alper, 1990) – for example, by learning to speak some of their language

– we would expect a nondysfunctional machine to do the same. Accordingly, the communion game may call upon machines to adapt to humans from different cultures.

In certain respects, our game resembles the total test of human performance proposed by Harnad (1989). Indeed, his test could be regarded as setting a goal beyond that of our game. At some future time, it might be desirable to update our test by requiring adaptive robots to produce and reply to visible bodily movements. To argue that this were necessary, however, would be to miss the point. This is because we believe it is useful to develop machines that take part in *language activity*. It is because talk depends on both words and biologically simpler types of patterning that we wish to revise the imitation game. Utterance meaning typically relies, not on winks and nods, but on verbal events. In conversations, as telephones and tape recordings show, visible behaviour play a relatively minor role. Since audible signals are usually sufficient for understanding conversation, there is no need to oblige our panel to consider visible behaviour (not to mention the role of touch or pheromones in human encounters). This is also a matter of not wishing to emphasise the building of human-like body suits. This would be a distraction from questions concerning language.

Approaches to Simulating Language Activity

Many computational advances have drawn inspiration from Turing's conjecture that machines could participate in language activity. However, perhaps for this very reason, the speech circuit model has remained unchallenged by computer scientists. As a consequence, and without exception, language has been modelled with respect to word-based forms. The result is that the phrase *speech and language processing* has acquired a strangely restricted sense, and the primary goals of the field are understood to be the design of machines that either transcribe and synthesise spoken prose or write and "interpret" text (see Huxor, 1994). Our communion game emphasises that the manipulation of word-based forms is only a part of conversation and that to participate machines will have to concert prosodic patterning across vocalisations.

Our game also emphasises that conversational events take place *between* individuals. Talking machines, like living organisms, must adapt to one another in a shared environment, and competition and cooperation become matters of strategy. Because adaptive coordination is observed both in vertebrates like birds and fish as well as in human infants as young as two months, investigation of how this co-behaviour is concerted provides an alternative starting point for research into language.

Symbol Manipulating Techniques: A Marriage of Convenience

Speech circuit models encourage engineers to think of utterances with respect to how word-based forms can be mapped onto one another. The marriage of speech circuit models and digital computers has resulted in programs that can take dictation from carefully spoken prose and conversely, read aloud in a neutral tone. Even recent attempts to produce "prosodically appropriate" synthetic speech rely on text-based syntactic and semantic analyses as well as prosodic norms that wash out the

characteristic ways in which individuals use their voices to get through to one another (Steedman, 1991; Prevost and Steedman, 1994). This marriage has also resulted in programs that produce text to pose as psychologists and their patients, UNIX consultants and Voyager II experts in keyboard-mediated conversations (Weizenbaum, 1965, 1976; Colby et al., 1972; Wilensky, 1983, pp. 151–156; Katz, 1990, respectively). Similar principles were applied in developing programs like SHRDLU that, in effect, map grammatical analyses onto simulated movements (Winograd, 1972). This approach has also been extended to the manipulation of real objects (Nilsson, 1984).

Among the disadvantages in developing computing technology in line with the speech circuit are that engineers take an abstract view of language, a view that invariably marginalises the role played by meaning. Following the younger Chomsky (1957, 1965), languages are usually defined with respect not to utterances but to abstract objects such as *sentences*. While Turing himself might have been unhappy with this move,[20] it is probably an inevitable consequence of the role assigned to teleprinters and the speech circuit in the early days of computing. Projects like those mentioned above are of interest primarily for their commercial applications to text-related problems.

By conceiving of these projects as inquiries into the use of "natural language", attention is drawn away from the hole in the bucket. It is not noted that they give no insight into what people mean as they speak. Engineers are encouraged to model utterances as if they were sentences accompanied by phonological, grammatical, and semantico-pragmatic representations. They are not encouraged to inquire into the point of what is said. Applied to *Daddy you're supposed to help me with my plane*, not only is it unclear what the abstractions would be, but it is also clear that they would be of little help in interpreting the utterance as the sister does. Her recognition that her brother is *battling to win her father's attention* depends less on what is said than on prosodic information systematically filtered out by the speech circuit.

Thus, in spite of the importance of computational techniques drawing on the speech circuit, these techniques have contributed little to our understanding of conversation. To borrow an analogy from Schieber (1994), current programs designed to play the imitation game are to talking as jumping on springs is to flying (p. 76). If we wish to come to grips with how people get through to each other in ordinary conversations, then to extend current models without rethinking our assumptions about language would simply be like building bigger springs. Thus, while of enormous practical value, the models described provide no insight into conversational activity. Even if they say something about the abstract aspects of meaning (semantics), they have nothing to say about meaning in ordinary senses of the word. This is because, in conversations, meaning is expressed not so much by words actually spoken as by how utterances are concerted between the individuals involved.

Architectures for Symbol Manipulating

We believe, on *a priori* grounds, that the case against speech circuit models is compelling. In spite of this, it is of interest to consider the kinds of goals arising when such models are adopted. Revealingly, to describe programs reflecting speech circuit models, a metaphor from the building trade is used: they are said to have *architectures*.

Underlying them is the assumption that we know what materials to employ and fit these to the machines at our disposal. The programmer's job, therefore, is to frame the task properly so as to provide for all eventualities. The processing of speech is decomposed into a long series of levels. A situation is usually modelled with explicit word-based representations (e.g. propositions) and so-called inferences are made on them to form "plans". Although standard architectures may use prosodic information – for example, to parse ambiguous sentences like *She gave her baby food* (see Waibel, 1988) – utterances are primarily treated as text. This can be pictured as in Fig. 2 below. As the diagram shows, according to this approach machines are to possess not only linguistic competence, but also the wherewithal for integrating output from semantico-pragmatic analysis with modules implementing synthesised "reasoning" and "planning".

It is, of course, legitimate to argue that specific models are important in computing and if these do not correspond to anything in human neurophysiology, this does not matter. Even on this view, it has to be conceded that architectures like the one just outlined often lead to systems that are hard to debug, slow in responding, and inflexible. Also, because inferences that appear reasonable in isolation may interact in ways that later produce incoherent conclusions, the programmer is required to anticipate a rule's often unforeseeable consequences.[21] Furthermore, these architectures are heavily reliant on their programming. This is because the machine's model of the world derives from the programmer's introspections and not from its own experience. It is unsurprising that such machines are often incapable of adapting to circumstances for which their programmers have not specifically prepared them.

Apart from these practical problems of implementation, other difficulties arise. It is highly controversial that even a single entirely grammatical sentence possesses

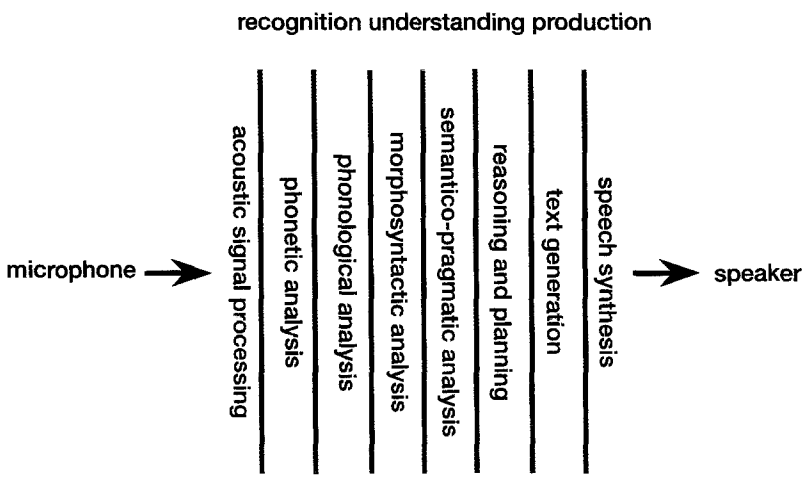

Fig. 2. A possible architecture for simulating conversation along traditional lines (see Klatt, 1977; Wheddon & Linggard, 1990). The speech-signal is analysed into text which in turn is analysed into a canonical form such as conceptual dependency (Schank & Abelson, 1977). This is done so that segments of speech having the same descriptive meaning can be represented equivalently. Formal reasoning and planning then act on these forms. If a reply is called for, the resulting forms must be converted into prose and then synthesized into speech.

specifiable phonetic, phonological, morphosyntactic, or semantico-pragmatic properties. Only a brave or foolish linguist would claim to be able to provide a *definitive* analysis of even an utterance as simple as that corresponding to *The president of France is bald.* Indeed, different formal descriptions could be developed at each of the levels.[22] Even if satisfactory analyses were provided, this task pales besides that of getting a machine to undertake human reasoning and planning. Not only is it unclear what is meant by asserting that people's utterances reflect reasoning and planning, but even if this issue were resolved (say, by a full understanding of neurophysiology) it is not obvious that this would facilitate programming.

Let us suppose, however, that machines were developed to synthesise reasoning and planning. Let us further imagine that such a machine were turned on and, after a few minutes of interaction, a human uttered: *A fly is swimming in my soup.* Depending on how it was spoken, this exclamation might well express any of a range of sentiments including horror and amusement. Since the machine would be unable to exploit these cues, it would have to rely on certain normative assumptions. It would almost certainly miss the point. What if it said, *I'm sorry. I know how upsetting that can be*? A machine that consistently strung together word-based forms as congruously as this might well win prizes at the Turing Test. But so what? If the human were enjoying watching the fly, which in all likelihood tones of voice would reveal, this response would be wholly inappropriate. Our imaginary computer, in other words, would be playing the game with a handicap. It would be condemned to *misinterpret* what was said because the structures it could access would be limited to those associated with the word-based form instantiated. If such machines were built, during ordinary conversations we would no more understand the point of what they said than they would grasp the force of our words.

Designing Adaptive Robots

The communion game focuses attention on biological issues. Rather as focusing on grammar raises questions about the mind, focusing on prosody raises questions about the body. To design machines that vocalise and modulate vocalisations functionally and with human-like precision is a matter of providing them with the ability to adapt to one another in real time. Accordingly, in urging that the imitation game be updated, we concentrate on questions that bring this into the open.

There are major contrasts between architectures like that described above and our design heuristic which focuses only on wordless prosody like that seen in infants (see Fig. 3). First, we aim not to develop a program but to design a type of robot. Second, we expect these robots to adapt to events taking place in a shared environment. Third, though some of what they do will lend itself to description in terms of traditional sense-model-plan-act architectures, the robots will have to integrate this strand of behaviour with simpler reactive behaviour. Finally, our robots will depend on audible word-based and phonetic patterning (and visual patterning) established over a set of encounters.

Turning to more specific matters, rather as computers transformed the notion of *response* into that of *output*, we expect adaptive robots to transform *input* into what we call *feature selection.* Like animals, our machines are imagined as living in *Merkwelten*[23] (perceptual worlds) whereby they not only detect features of their

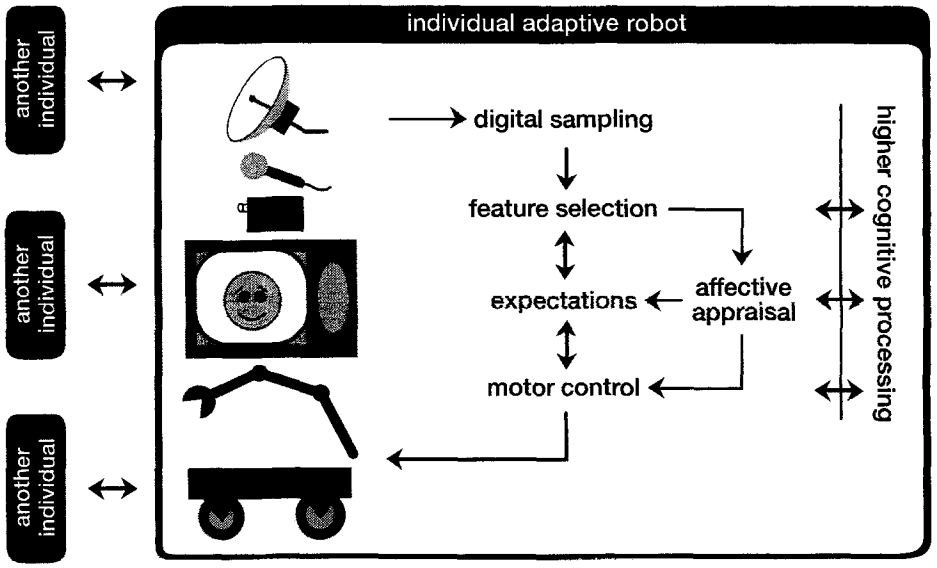

Fig. 3. An heuristic for the design of robots that might participate in the communion game. A robot digitally samples readings from its sensors and selects features of recent readings. These features are used to recall experiences likely to be relevant to the robot's current circumstances. This is accomplished by giving more weight to features which, in similar situations, have proved useful in predicting effects achieved by the robot's actions. From its experiences, and its affective appraisals, the robot develops expectations about probable consequences of its actions, and these serve as a constantly developing model of the world. Motor control takes appropriate action based on the robot's currently elicited expectations.

environment but also adapt to notice new relevant aspects of it. What they listen to and look at will reflect not just what is objectively there but also whom they are interacting with and their own inner states. This aspect of the model is motivated by observation. People do not respond to word-based forms but construe utterances; when words are spoken, different people select different features as worthy of reply.

Since what is meant depends not only on what is said but also on adaptive vocal coordination between individuals, we depart substantially from traditional sense-model-plan-act architectures. This is indicated by a double-headed arrow linking feature selection with *expectations*. Because of our interest in what people do, following Martin and Bateson (1986) behaviour is defined in terms of action *and* reaction. Thus, even if in this context we are playing down *higher cognitive processes*, this is not because we underrate their importance. Rather, it is because we believe that investigating prosodic aspects of language will provide insights into the higher processes.

We borrow from observations about animal behaviour to show that *motor control* does not simply reflect output from posited higher cognitive processes. Our robots, like living organisms, do not merely act. Rather, they engage in *activity* fully integrated with what is happening both internally and in the environment. Thus, in the cited conversational exchange, when *B* "replies" by saying *I have the put the glue*, the word-based forms can be regarded as showing that she follows the dictates of higher cognitive processes. Simultaneously, however, the timing and way of speaking indicate

a reactive response to her brother's ongoing activity. One cannot do justice to the event taking place as this utterance is spoken without considering how the sister's activity is integrated with her brother's.

To capture relations between an individual's *Merkwelt* and what that individual is doing, the figure employs a column of icons to represent sensors and actuators used to concert behaviour between robots. It is to account for this complexity that we introduce the notion of expectations (see Sommerhoff and MacDorman, 1994, section 5), influenced simultaneously by current *affective appraisals* and higher cognitive processes.

Getting Down to Brass Tacks

It may be easier for the reader to envisage a working inner loop after we have outlined one possible approach to its implementation. Readings from all the robot's sensors are sampled at discrete intervals (e.g. 10 ms), digitised, and temporarily help in sensory buffers. To simplify matters we can consider two aspects of the acoustic signal only: the fundamental frequency and amplitude of the robot and its current conversation partners' recent vocalisations. These aspects constitute its perceived situation. The robot will try an action (e.g. a vocalisation or silence) in a given situation. Once it has detected the consequences, it stores the action together with the original and new situation. The robot develops expectations about the consequences of its possible actions from these experiences.

Initial affective appraisals are made in certain situations such as those in which the robot has attained or failed to attain a reward. In such cases the robot can update its experiences leading towards the reward (or undesirable situation) with either information about its distance from it (see Michie and Chambers, 1968) or discounted rewards (see Sutton, 1988; Watkins and Dayan, 1992). Eventually the robot can determine whether an action is likely to lead toward or away from a reward.

A sensible way for a robot to learn to predict the probable consequences of an action (e.g. a vocalisation) is by remembering what happened when it did the same thing in similar situations (e.g. with the same partner). In this way, it can bring its experiences to bear on its current behaviour. The important issue here is determining which past situations are similar to the present one and, indeed, what *similar* should mean in the present circumstances (and relationships). Situations should be distinguished from one another in such a manner that, given an action, those more likely to lead to similar outcomes are identified as more similar than those less likely to do so.

A particular situation can be viewed as a point in multidimensional phase space (see Clocksin and Moore, 1989; Omohundro, 1990) whose location in determined by measures of the fluctuating frequency and amplitude of the robot and its partners' recent vocalisations. Good prediction can be achieved by finding for the point corresponding to the immediately preceding patterning a set of weights such that a locally weighted distance function can interpolate from its nearest neighbours the most likely outcome of an action. Dimensions that are more relevant in a certain region of space are accentuated, and those that are less relevant diminished. Finding and exploiting the best generalisations from experiences in this manner is tantamount to feature selection. The expectations the robot develops as a result serve as its model of the world.

Enough has been said to show that our communion game redefines Turing's goal of getting machines to participate in language activity. Above all, it focuses the mind not on hypothetical structures, but on how to model complex behaviour taking place between organisms. This in itself should be sufficient to shift the emphasis from programming computers to developing adpative robots. It is worthy of note that Turing himself anticipated this move when he ended his paper by discussing learning machines and suggested that they could be given sense organs and then taught English (p. 460).

Since our game is designed to promote the development of machines able to participate in language, it does not presuppose that we know what languages are like. It treats this as a question open to empirical as well as theoretical investigation. Machines do not need to deal with the whole of the language at once. Different kinds of technology and design will be useful in coming to terms with different aspects of it.

Indeed, we believe that the place to start is by simulating machines that adapt to what each other is doing. Not only does this capture what typically happens in vertebrate communication, but it is also necessary to play the communion game. Thus, a research program might start with interacting agents that only exist in a computer simulation. These would learn to communicate with each other to achieve specifiable ends. Neural networks, reinforcement learning, genetic algorithms, phase space approaches (like the one just outlined) or some combination thereof may be experimented with in these early systems. If these systems proved satisfactory, the agents could later be embodied in robots active in real or artificial worlds. Eventually, these robots would be required to learn to interact adaptively with humans. This, no doubt, would be difficult. Wittgenstein noted that "if a lion could talk, we could not understand him" (1958, p. 223). However, we are fairly confident that if robots (or lions) could be evolved that could learn to talk to each other, they could also learn to talk with us.

Once it were possible to start building actual machines, it would be time to focus on prosodic coordination between agents. In short, this step would involve developing robots to implement the inner loop shown on the diagram above. At this stage attention would also be focused on prelogical aspects of language organised according to principles similar to those observed in other animals. This can, we believe, be readily justified. Although there are evolutionary and developmental grounds for recognising that word-based aspects of language cannot be separated from prosody, prosody is separable from words: babies do not exploit words as they interact vocally with their caretakers. By deciding to focus on the inner loop of the diagram one might start with prosodic modelling. Later, and especially if these machines developed what a human observer identified as arbitrary symbols, it would be time to introduce the communion game.

Concluding Remarks

Conversations are crucial to intelligence. If you asked the man in the street whether a robot that collects litter or sorts widgets were showing this quality, he might well deny it. But if you asked him whether a robot that could hold its own in a conversation

were intelligent, he would be likely to say *yes* – even if that conversation were only with his three-year-old daughter.

Brooks (1991a,b) challenges this commonsensical attitude and argues that if we are to understand intelligence, we must begin where intelligence first appeared: we must begin with the sensorimotor coordination of simple creatures. From this point of view, were we to begin with conversation, something that has appeared comparatively late in evolutionary time, we would be starting our journey from too near the finish line. Brooks believes that this mistake has distorted the enterprise of simulating intelligence. Were conversations literally *composed of* word-based forms, we might have agreed with him. However, since they are, in fact, vocal events consisting of words that depend on complex sensorimotor coordination, we believe it may be better to approach the problem from the developmental starting blocks. Further, since talk depends not only on what individuals think separately but also on how they *publicly* attune their ideas, we have reason to believe progress can be made especially in modelling prosodic events. Finally, the parallels between the aspects of speech and the closely coordinated behaviour of simpler creatures are unmistakable. We believe that designing a social group of adaptive robots is a first step in designing robots that develop their own language. If machines can develop their own languages, we shall have hopes of teaching them ours.

Since the early days, many computer scientists have believed that by modelling intelligence we can gain insight into how people think. These insights have failed to materialise partly because researchers have not come to terms with the fact that we can only simulate behaviour. In this paper, we set out to remedy this shortcoming. We are not calling into account the spirit of the Turing Test which, insofar as it encourages empirical work, is unrivalled. What we are calling into account is both the machinery with which Turing expected to carry through his project and the artificial and unilluminating model of conversation implicit in his game. By paying attention to the part played by prosodic features of speech during ordinary talk, the inadequacy of speech circuit models is highlighted. Having done this, like Turing, we have restated the goal as an engineering problem.

Our communion game is designed, not to resolve philosophical puzzles concerning the nature of intelligence, but to stimulate the development of machines that talk. We believe that we have identified a goal of considerable theoretical, practical, and commercial importance and are only surprised that it has not previously been specified. Talking machines would impress not just computer scientists but also the people down the road.

References

Allen, W.S. (1973). *Accent and rhythm prosodic, features of Latin and Greek: A study in theory and reconstruction.* Cambridge: Cambridge University Press.

Alper, G. (1990). A psychoanalyst takes the Turing test. *Psychoanalytic Review, 77*(1), 59–68.

Atkinson, J.M. & Heritage, J., Eds. (1984). *Structures of social action: Studies in conversation analysis.* Cambridge: Cambridge University Press.

Auer, P. and di Luzio, A., Eds. (1992). *The contextualization of language.* Amsterdam: John Benjamens.

Bateson, G. (1979). *Mind and nature: A necessary unity.* New York: Ballantine.

Bateson, P.P.G. (1988). Biological evolution of cooperation and trust. In D. Gambetta (Ed.), *Trust: Making and breaking of cooperative relations.* Oxford: Blackwell.

Bieri, P. (1988). Thinking machines: Some reflections on the Turing test. *Poetics Today, 9*(1), 163–186.
Block, N. (1978). Troubles with functionalism. In C.W. Savage (Ed.), *Perception and cognition: Issues in the foundations of psychology. Minnesota studies in the philosophy of science* (Vol. 9, pp. 261–325). Minneapolis: University of Minnesota Press.
Block, N. (1981). Psychologism and behaviorism. *The Philosophical Review, 90*(1), 5–43.
Bloomfield, L. (1933). *Language*. New York: Henry Holt.
Brooks, R.A. (1991a). Intelligence without reason. In *IJCAI–91: Proceedings of the Twelfth International Conference on Artificial Intelligence,* Sydney, Australia (Vol. 1), pp. 569–595. San Mateo, CA: Morgan Kaufmann.
Brooks, R.A. (1991b). Intelligence without representation. *Artificial Intelligence, 47*, 139–159.
Brown, E.D. (1979). The song of the common crow, *Corvus brachyrhynchos*. Master's thesis at University of Maryland, College Park.
Carling, C. & Moore, T. (1982). *Language understanding: Towards a post-Chomskyan linguistics*. New York: St. Martin's Press.
Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Chomsky, N. (1986). *Knowledge and language: Its nature, origin, and use*. New York: Praeger.
Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures*. Cambridge, MA: MIT Press.
Clocksin, W.F., & Moore, A.M. (1989). Experiments in adaptive state-space robotics. In *AISB89: Proceedings of the Seventh Conference of the Society for the Study of Artificial Intelligence and Simulation of Behaviour,* Brighton, UK, pp. 115–125. San Mateo, CA: Morgan Kaufmann.
Colby, K.M., Hilf, F.D., Weber, S., Kraemer, H.C. (1972). Turing-like indistinguishability tests for the validation of a computer simulation of paranoid processes. *Artificial Intelligence, 3*, 199–221.
Cowley, S. J. (1993). *The place of prosody in Italian conversations*. Unpublished doctoral dissertation, University of Cambridge, Cambridge, UK.
Cowley, S.J. (1994). Conversational functions of rhythmical patterning — A behavioural perspective. *Language & Communication, 14*(4), 353–376.
Cowley, S.J. (1996). Conversation, coordination, and vertebrate communication. *Semiotica*.
Crystal, D. (1969) . *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.
Dreyfus, H.L. (1992). *What computers still can't do: A critique of artificial reason*. Cambridge, MA: MIT Press.
Farabaugh, S. M. (1982). The ecological and social significance of duetting. In D. Kroodsma, E.H. Miller, and H. Ouelett (Eds.), *Acoustic communication in birds: Song learning and its consequences* (Vol. 2, pp. 85–124). London: Academic Press.
Fodor, Jerry, (1975) *The language of thought*. New York: Cromwell.
French, R.M. (1990). Subcognition and the limits of the Turing test. *Mind, 99*(393), 53–65.
Giles, H. & Coupland, N. (1991). *Language: Contexts and consequences*. Milton Keynes: Open University Press.
Gödel, K. (1931). Uber formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik, 38*, 173–198.
Gumperz, J.J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.
Gunderson, K. (1971). *Mentality and machines*. New York: Doubleday.
Halliday, M.A.K. (1978). Meaning of the construction of reality in early childhood. In H. L. Pick, Jr., & E. Salzman (Eds.), *Modes of perceiving and processing information* (pp. 67–96). Hillsdale, NJ: Erbaum.
Halliday, M.A.K. (1979). One child's protolanguage. In M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication* (pp. 171–190). London: Cambridge University Press.
Harnad, S. (1989). Minds, machines and Searle. *Journal of Experimental and Theoretical Artificial Intelligence, 1*, 5–25.
Harnad, S. (1990). The symbol grounding problem. *Physica D, 42*, 335–346.
Harré, R. & Gillett, G. (1994). *The discursive mind*. Thousand Oaks, CA: Sage Publications.
Harris, R. (1981). *The language myth*. London: Duckworth.
Harris, R. (1987). *The language machine*. London: Duckworth.
Hart, J. 't, Collier, R. & Cohen, A. (1990). *A perceptual study of intonation: An experimental approach to speech melody*. Cambridge: Cambridge University Press.
Hauser, L. (1993). Reaping the whirlwind: Reply to Harnad's "Other Bodies, Other Minds." *Minds and Machines, 3*(2), 219–237.
Hobbs, A.G., Yeomanson, E.W., & Gee, A.C. (1973). *Teleprinter Handbook*. Hertfordshire, UK: Radio Society of Great Britain.
Hultzén, L.S. (1964). Grammatical intonation. In D. Abercrombie, D.B. Fry, P.A. D. MacCarthy, N.C. Scott, J.L.M. Trim (Eds.), *In honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday* (pp. 85–95). London: Longmans.

Huxor, A.P. (1994). *Artificial intelligence as writing: Knowledge-based hypertext systems as medium for communication.* Unpublished doctoral dissertation, Middlesex University, Middlesex, UK.
Jacquette, D. (1993). A Turing test conversation. *Philosophy, 68*(264), 231–233.
Karelis, C. (1986). Reflections on the Turing test. *Journal for the Theory of Social Behavior, 16*(2), 161–172.
Katz, B. (1990). Using English for indexing and retrieving. In P.H. Winston and S.A. Shellard (Eds.), *Artificial Intelligence at MIT: Expanding Frontiers.* Cambridge, MA: MIT Press.
Klatt, D.H. (1977). Review of the ARPA speech understanding project. *JASA, 62*(6), 1345–1366.
Krebs, J.R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs and N.B. Davies (Eds.), *Behavioral ecology: An evolutionary approach* (2nd ed., pp. 380–402). Oxford, UK: Blackwell Scientific.
Laver, J. (1980). *The phonetic description of voice quality.* Cambridge, UK: Cambridge University Press.
Laver, J. (1993). *Principles of phonetics.* Cambridge, UK: Cambridge University Press.
Levinson, S.C. (1995). Interactional biases in human thinking. In E.N. Goody (Ed.), *Social intelligence and interaction.* Cambridge: Cambridge University Press.
Lenat, D.B., & Guha, R.V. (1990). *Building large knowledge-based systems: Representation and inference in the Cyc project.* Reading, MA: Addison-Wesley.
Locke, J. (1993). *The child's path to spoken language.* Cambridge, MA: Harvard University Press.
Lowerre, T., & Reddy, D.R. (1980). The Harpy speech understanding system. In W.A. Lea (Ed.), *Trends in speech recognition* (pp. 340–360). Englewood Cliffs, NJ: Prentice-Hall.
Lucas, J.R. (1961). Minds, machines and Gödel. *Philosophy, 36,* 112–127.
Lyons, J. (1977). *Semantics* (Vols. 1–2). Cambridge, UK: Cambridge University Press.
Malinowski, B. (1923). The problem of meaning in primitive languages. In C.K. Ogden and I.A. Richards, *The meaning of meaning: A study of the influence of language upon thought and the science of symbolism* (pp. 451–510). London: Kegan Paul, Trench, Trubner & Co.
Martin, P.R., & Bateson, P. (1986). *Measuring behavior: An introductory guide.* Cambridge, UK: Cambridge University Press.
Matthews, P.H. (1993). *Grammatical theory in the United States from Bloomfield to Chomsky.* Cambridge, UK: Cambridge University Press.
McDermott, D. (1993). Book review: Building large knowledge-based systems: Representation and inference in the Cyc project, D.B. Lenat & R.V. Guha. *Artificial Intelligence, 61*(1), 53–63.
Mey, J.L. (1993). *Pragmatics.* Oxford: Basil Blackwell.
Michie, D. (1993). Turing's test and conscious thought. *Artificial Intelligence, 60*(1), 1–22.
Michie, D., & Chambers, R. (1968). Boxes: an experiment in adaptive control. In E. Dale and D. Michie (Eds.), *Machine Intelligence* (Vol. 2, pp. 137–152). Edinburgh: Oliver & Boyd.
Nilsson, N.J. (1984). Shakey the robot. Technical Note 323, SRI AI Center, Menlo Park, CA.
Ogden, C.K., & Richards, I.A. (1923). *The meaning of meaning: A study of the influence of language upon thought and the science of symbolism.* London: Kegan Paul, Trench, Trubner & Co.
Omohundro, S.M. (1990). Geometric learning algorithms. *Physica D, 42*(1–3), 307–321.
Penrose, R. (1989). *The emperor's new mind.* Oxford, UK: Oxford University Press.
Prevost, S. & Steedman, M. (1994). Specifying intonation from context for speech synthesis. *Speech and communication, 15,* 139–153.
Saussure, F. de (1916). *Course de linguistique générale.* Paris: Payot. English Translation [1959], *Course in general linguistics.* London: Peter Owen.
Schank, R.G., & Ableson, R.P. (1977). *Scripts, Goals, Plans and Understanding.* Hillsdale, NJ: Erlbaum.
Searle, J.R. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences, 3,* 417–457.
Shannon, C.E. and Weaver, W. (1949). *The mathematical theory of communication.* Urbana, IL: University of Illinois Press.
Shanon, B. (1989). A simple comment regarding the Turing test. *Journal for the Theory of Social Behavior 19*(2), 249–259.
Shieber, S.M. (1994). Lessons from a restricted Turing test. *Communications of the ACM, 37*(6), 70–78.
Sims, K. (1994). Evolving 3D morphology and behavior by competition. *Artificial Life IV Proceedings,* pp. 28–39. Cambridge: MIT Press.
Slezak, P. (1982). Gödel's theorem and the mind. *British Journal for the Philosophy of Science, 33,* 41–52.
Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11*(1), 1–23.
Sommerhoff, G., & MacDorman, K. F. (1994). An account of consciousness in physical and functional terms: A target for research in the neurosciences. *Integrative Physiological and Behavioral Science, 29*(2), 151–181.
Sperber, D. & Wilson, D. (1986). *Relevance: Communication and cognition.* Oxford: Basil Blackwell.
Steedman, M. (1991). Structure and intonation. *Language, 68,* 260–296.
Strawson, P.F. (1950). On referring. *Mind, 59,* 320–344.
Sutton, R.S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning, 3*(1), 9–44.

Thorpe, W.H. (1972). Duetting and antiphonal song in birds: Its extent and significance, *Behavior*, Suppl. *18*, pp. 1–197.
Trevarthen, C. (1979). Communication and co-operation in early infancy: A description of primary intersubjectivity. In M. Bullowa (Ed.), *Before speech: The beginning of interpersonal communication* (pp. 321–348). Cambridge: Cambridge University Press.
Trevarthen, C. (1986). Sharing makes sense: Intersubjectivity and the making of an infant's meaning. In R. Steele and T. Threadgold (Eds.), *Language Topics: Essays in honour of M. Halliday*, (Vol. 1, pp. 177–200). Amsterdam: J. Benjamens.
Turing, A. (1950). Computing machinery and intelligence. *Mind, 59*, 433–460.
Üxküll, J. J., Baron von (1921). *Umwelt und Innenwelt der Tiere*. Berlin: Springer.
Waibel, A. (1988). *Prosody and speech recognition*. London: Pitman.
Watkins, C.J. C.H., & Dayan, P. (1992). Q-learning. *Machine Learning, 8*(3–4), 279–292.
Weizenbaum, J. (1965). ELIZA—a computer program for the study of natural language communication between man and machine. *Communications of the Association for Computing Machinery, 9*, 36–45.
Weizenbaum, J. (1976). *Computing power and human reason: From judgment to calculation*. San Francisco: W. H. Freeman.
Wheddon, C., & Linggard, R. (1990). *Speech and language processing*. London: Chapman & Hall.
Wilensky, R. (1983). *Planning and understanding: A computational approach to human reasoning*. Reading, MA: Addison-Wesley.
Winograd, T. & Flores, F. (1986). *Understanding computers and cognition: A new foundation for design*. Norwood, NJ: Ablex.
Winograd, T. (1972). *Understanding natural language*. New York: Academic Press.
Wittgenstein, L. (1958). *Philosophical investigations*. Oxford, UK: Basil Blackwell.

Notes

1 These are the terms adopted from Laver, 1993. In what follows aspects of speech are treated as being *prosodic* in that, without being strictly associated with the phonological forms used, they reflect how a speaker modifies his voice as he speaks. In Hultzen's (1964) phrase they represent the "residue of utterance" and are reflected especially in modulations of pitch, loudness, and rapidity. (For an historical discussion of the scope of προσωδια and prosody, see Allen, 1973; for discussion of its auditory and acoustic correlates, see Crystal, 1969). Voice quality is defined by Laver (1980) as "the characteristic auditory colouring of an individual speaker's voice" (p. 1).
2 Observations that, in some ways, are comparable to those reported here are to be found in the work of Auer and his colleagues (Auer and di Luzio, 1992).
3 The schools of linguistics in which they are given most prominence are in conversation analysis (see Atkinson & Heritage, 1982), accommodation theory (see Giles & Coupland, 1991) and in interactional sociolinguistics (see Gumperz, 1982; Auer & di Luzio, 1992). None of these subdisciplines, however, focus on how prosodic effects influence what, in an ordinary sense, utterances mean. The work of 't Hart, Collier, and Cohen (1990) is directed at addressing this issue. For a more linguistic approach see Crystal (1969).
5 Different speakers reading aloud from the Wall Street Journal is a typical example of a training set. Animated and spontaneous speech has thus far been considered too difficult.
6 This argument is not original. It is put particularly forcefully by Harris (1987). Many others have argued a closely related view: that language cannot be described in relation to what is often called the code view (for an elegant exposition, see Sperber and Wilson, 1986).
7 These excellent and diverse publications do not merely adopt speech circuit models. They adapt them to their own purposes: Ogden and Richards propose such a model as the basis for "scientific or symbolic" communication; Bloomfield finds that the model accords with his goal of systematically distinguishing "linguistic forms" from speech; finally, Shannon and Weaver adopt a related model because it provides the basis for an explicit mathematical theory with engineering applications.
8 Or, in Saussure's terms *signes* (signs).
9 Historically, the most important of these are *langue* (Saussure, 1916), speech habits (Bloomfield, 1933), competence (Chomsky, 1965) and I language (Chomsky, 1986, 1988).
10 Turing (1950) introduced his imitation game by asking us to envisage a game played by a man and a woman. The man's objective is to convince an interrogator that he is the woman while she attempts the same. The contestants communicate with the interrogator in written form, ideally across a teleprinter, thus concealing their identities. "We now ask the question," Turing writes, "'What will happen when a machine takes the part of [the man] in this game?' Will the interrogator decide wrongly as often when the game is played like this as he does when the game is played between a man and a woman?" (p. 434)

It has often been mistakenly assumed that the game required the computer to simulate, not a woman's conversation, but a human's. Turing himself appears to make the confusion at one point: "it will naturally be assumed that the best strategy is to try to provide answers that would naturally be given by a man" (p. 435).

11 A teleprinter is a device used to punch or print out a machine readable tape. For further details on teleprinters with illustrations see Hobbs, Yeomanson and Gee (1973). Teleprinters are called teletypes in the U.S. The main updating of the Turing Test over the past forty odd years is the substitution of the network terminal for the teleprinter. This, however, does not affect the arguments presented in this paper.

12 In this respect generative grammarians remain faithful to the central tenets of structuralism. (For exposition of these parallels see Matthews, 1993.)

13 The reason for this muddle lies in the fact that literate adults can interpret texts as if they were sequences of symbol strings. This leads one into the temptation of supposing that, for the human as for the computer, they *are* symbol strings. As we have seen, during ordinary conversation, this text-based picture is indisputably false.

14 Shanon (1989) notes that the imitation game purports to demonstrate what is covertly assumed: namely that insofar as machines can manipulate symbols they are, in this respect, like people.

15 To avoid confusion, drawing on the Shorter Oxford English Dictionary (1933, p. 1897, 2nd entry), we construe the term to assert that something has the "same external features" as that which is simulated.

16 In describing our game as a communion game, we allude to Malinowski's "phatic communion" (1923). The anthropologist noted that in simple types of conversation or, in his terms, in "primitive uses of language," speech "functions as a link in concerted human activity, as a piece of human behavior. It is a mode of action and not an instrument of reflection" (p. 474). It is not accidental that Gregory Bateson echoes this with his emphasis on communion (1979).

17 It is noteworthy that contemporary voice recognition systems attempt to factor out nonsegmental aspects of utterances such as rhythm, pitch, prosody, and voice quality (Waibel, 1988). Were the engineer to come to grips with them, perhaps by anticipating them instead of treating them as noise, the task of recognizing word-based aspects of utterances would be easier.

18 For current approaches to the Turing Test, the structures used to compose replies are entered by a programmer rather than learnt. The most ambitious example of a system capable of drawing inferences from programmed 'knowledge' is Cyc (Lenat & Guha, 1990). This approach, however, has come under much criticism (e.g., McDermott, 1993), and Turing himself remarks on its shortcomings, arguing instead for machines that learn from their own experiences and especially from being taught (p. 455–456).

19 We include here questions designed to disclose the machine's inability to recognize that a mathematical proposition is true but unprovable (Jacquette, 1993; also see Gödel, 1931; Turing, 1950; Lucas, 1961; Slezak, 1982; Penrose, 1989), questions designed to reveal 'subcognitive' differences by associative priming and rating games (French, 1990), and questions designed to trick the machine into articulating more knowledge about its own processes of reasoning than a human could (Michie, 1993).

20 Turing did not suppose that individuals mapped language onto semantic representations. He appears to have accepted that language is normative, that some utterances are corrigible, and that it is mainly learned at school (see especially p. 456).

21 Referred to as the *frame problem*, this has been a longtime stumbling block for progress in symbolic AI, see Harnad (1990, p. 339).

22 Actual systems may have more or fewer levels than shown in Figure 2. For example, Harpy combined syntactic, semantic, and pragmatic analysis into a single augmented transition network (Lowerre & Reddy, 1980). This suggests that the standard distinctions linguists and engineers make between speech processing level are not natural but motivated by convention and explanatory convenience.

23 The way an adaptive machine will model its world necessarily has much to do with the experience it has accrued through using its particular sensors and actuators in coming to terms with its *Merkwelt*, that is, its own constantly updated individual perceptual world (see Brooks 1991b; Üxküll, 1921).

24 In vocal communication the best examples of close coordination currently belong to birds. Thus, for example, Thorpe (1972) describes how a pair of bou-bou shrikes learned to duet; Brown (1979) reports that North American crows vary their calls depending on whether they are interacting with kin, mates or strangers and Farabaugh (1982) reports that buff breasted wrens time their duets so finely that a human observer can tell how far apart they are in the forest.

Correspondence and offprint requests to: Karl MacDorman, Computer Laboratory, New Museums Site, Pembroke Street, Cambridge CB2 3QG, UK.