Introduction

Programs for the simulation of human chat, such as ELIZA, DOCTOR, and PARRY, confine themselves to putting human conversants at their ease with polite and sometimes flattering questions, while deflecting counter-questions. Because these programs fail the main goal of facts-exchange, their historical significance has been overlooked. It is as though bronze age pioneers, having fashioned timber wheels and axles, had set people-carrying contraptions to run down gentle slopes. The false appearance of having invented automobiles, once debunked, could then obscure the true achievement (wheeled transport). ELIZA supplied demonstrably functional wheels for human chat. But the lack of engine, steering and guidance obscured the important lesson that however humdrum an automaton's homeostatic goals", they may yet constitute preconditions of later "achievement goals".

Results reported here are not yet sufficiently solid to meet normal criteria of publication. But initial observations with a prototype museum-guide speak so unambiguously to this and some other design issues that we have used their admittedly anecdotal testimony as a focus for a historical and methodological review. Very recently, results have begun to arrive weekly from a newly installed field site. These will be summarised in approximately yearly reports following on from this paper. The present immediate observations have been

(1) that the *achievement* goals of automated conversation (in this case information-provision to visitors to a science museum) can only be attained *via* continuous maintenance of rapport in the user towards the agent;

(2) that *homeostatic* goals of rapport maintenance are satisfiable by appropriate switching between informative and chat modes through context control mechanisms described by one of us elsewhere in this volume (Sammut, 2001).

The terms achievement and homeostatic correspond to the "better goals" and "holding goals" of Advice Language (Bratko and Michie, 1980), and to the "better" and "worse" goals of Barbara Huberman's (1968) classic study of programming the play of the elementary chess end-games. This is no accident. Conversation is a two-person game. In beta-testing it is zero sum; in the field the aim is win-win.

Our route map has been Turing's "child machine" three-stage approach. After two years we have essentially completed Stage 1 (see later) and are taking the first steps into Stage 2. But this paper is not about the design and implementation of the museum-guide agent, nor about the architecture of the scripting engine, nor about its *PatternScript* scripting language, nor about the new generic art of agent-scripting. Adequate coverage of these topics would constitute a full-semester graduate course. Rather, it reviews the rather wide range of behavioural features common to diverse applications areas. The link that binds new history with old are those intrinsic to any child-machine approach whatsoever. Recall that Stage 3 involves handing over educable and trainable systems to professional teachers and coaches. A minimal level of rapport-inducing and socially predictable behaviour must be guaranteed to people accustomed to demand these characteristics from their human pupils.

Statement of commercial requirements

To meet the new requirements, the following is a minimal set of what must be simulated.

Real chat utterances are concerned with associative exchange of mental images. They are constrained by contextual relevance rather than by logical or linguistic laws. Time-bounds do not allow real-time construction of reasoned arguments, but only the retrieval of stock lines and rebuttals, assembled Lego-like on the fly.

A human agent has a place of birth, age, sex, nationality, job, family, friends, partners, hobbies etc., in short a "profile". Included in the profile is a consistent personality, which emerges from expression of likes, dislikes, pet theories, humour, stock arguments, superstions, hopes, fears, aspirations etc. On meeting again with the same conversational partner, a human agent is expected to recall not only the profile, but also the gist of previous chats, as well as what has passed in the present conversation so far.

A human agent typically has at each stage a main *goal*, of fact provision, fact elicitation, wooing, selling, "conning" etc. A human agent also remains ever-ready to maintain or re-establish rapport by switching from goal mode to chat mode and back. Implementation of this last feature in a scriptable conversational agent will be illustrated.

State of the art

Commercial programs are surfacing that can find their way, partly by what is ordinarily called bluff, through interactive chat sessions in what passes for colloquial English. In the United States of America the first patent was granted in the summer of 2001 to the company NativeMinds, formerly known as Neuromedia. One of its two founders, Scott Benson, co-authored a paper by Nils Nilsson in Volume 14 of this *Machine Intelligence* series. As Appendix 1 a press release is reproduced which gives a rather vivid sketch of the nature of the new commercial art. To the applications described there, the following may be added.

1. Question-answering guides at trade shows, conferences, exhibitions, museums, theme parks, palaces, archaelogical sites, festivals and the like.

2. Web-based wizards for e-commerce that build incrementally assembled profiles of the individual tastes and foibles of each individual customer.

3. Alternatives to questionnaires for job-seekers, hospital patients, applicants for permits and memberships, targets of market research, and human subjects of psychological experiments.

4. Tutors in English as a second language. There is an acknowledged need to enable learners to practise *conversational skills* in augmentation of existing Computer-Aided Language Learning programs.

The example developed by Claude Sammut and myself for the "Cyberworld" exhibition in Sydney Australia belongs to category 1. We judge it to be second only to category 4 in the demands made on developers.

Historical roots

The philosopher of mind Daniel Dennett (2001) regards Turing's original "imitation game" as more of a conversation-stopper for philosophers than anything else. In this I am entirely with him.

The *weak form* presented in the 1950 paper is generally known as the Turing Test. It allows a wide latitude of failure on the machine's part. To pass, the candidate need only cause the examiners to make the wrong identification, as between human and machine, in a mere 30 per cent of all pairs of conversations. Only five minutes are allowed for the entire man-machine conversation. Turing's original specification had a human interrogator communicating by remote typewriter link with two respondents, one a human and one a machine.

> I believe that in about fifty years' time it will be possible to programme computers, with a storage capacity of about 10^9, to make them play the imitation game so well that an average interrogator will not have more than 70 per cent chance of making the right identification [as between human and machine] after five minutes of questioning.

Dennett's view is re-inforced by an account I had from Turing's friend, the logician Robin Gandy. The two friends extracted much mischievous enjoyment from Turing's reading aloud the various arguments and refutations as he went along with his draft.

Turing would have failed the Turing Test

Note that the Test as formulated addresses the *humanness* of the respondent's thinking rather than its *level*. Had Turing covertly substituted himself for the machine in such a test, examiners would undoubtedly have picked him out as being a machine. A distinguishing personal oddity of Turing's was his exclusive absorption in the literal intellectual content of spoken discourse. His disinclination, or inability, to respond to anything in the least "chatty" would leave an honest examiner with little alternative but to conclude: "this one cannot possibly be the human; hence the other candidate must be. So *this* one must be the machine!"

Experimental findings are presented to the effect that chat-free conversation is not only generally perceived as less than human, but also as boring. The concluding reference to "banter" in Appendix 1 suggests that NativeMinds have come to a similar conclusion. It seems that for purposes of discourse we must refine the aim of automated "human-level intelligence" by requiring in addition that the user perceive the machine's intelligence as being of human type. A client bored is a client lost.

Weak form of the game obsoleted

Turing himself believed that beyond the relatively undemanding scenario of his Test, the capacity for deep and sustained thought would ultimately be engineered. But this was not the issue which his 1950 imitation game sought to settle. Rather, the quoted passage considers the time-scale required to decide in a positive sense the lesser and *purely philosophical* question: what circumstances would oblige one to concede a machine's claim to think at all?

In the 1950 *Mind* paper (p.14) Turing remarked:

> I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.

When terms are left undefined, meanings undergo subtle change over time. Before his projected 50 years were up, words like "think", and "intelligent" were already freely applied to an ever-widening range of computational appliances, even though none came anywhere near to success at even the weak form of the imitation game.

In today's statements of engineering requirements and in diagnostics we encounter the language not only of thought but also of intention, and even of conscious awareness. The following exchange is abridged from the diagnostics section of a popular British computing magazine, *What Palmtop and Handheld PC*, June 2000. I have underlined certain words, placing them within square brackets to draw attention to their anthropomorphic connotations of *purpose, awareness*, and *perception.*

> AILMENT: I recently purchased a Palm V and a Palm portable keyboard. But whenever I plug the Palm into the keyboard it [attempts] to HotSync *via* the direct serial connection. If I cancel the attempted Hot Sync and go into the Memo Pad and try to type, every time I hit a key it [tries] to HotSync. What am I doing wrong?

> TREATMENT: The most logical solution is that your Palm V is [not aware that] the keyboard is present. You will need to install or reinstall the drivers that came supplied with the keyboard and to make sure that it is enabled. This will stop your Palm V [attempting] to HotSync with the keyboard and to [recognise] it as a device in its own right.

Modern information technology follows medical practice in applying terms such as "aware" and "intelligence" provided only that a system responds in certain ways to certain kinds of input. A clinical psychologist administering intelligence tests does not worry that in obtaining a high score the subject might have been merely pretending to think. If, though, the subject is then found to have been a robot, a quandary arises similar to that faced by Gary Kasparov, the then reigning World Chess Champion, during his defeat by the chess machine Deep Blue. His surprise at the chess intelligence which he perceived in certain of his opponent's moves led him to level a strange charge against the Deep Blue team. Somehow they must have made this precious human quality accessible to the machine in some manner that could be construed as violating its "free-standing" status. The chess world as a whole preferred to credit Deep Blue with having its own kind of chess intelligence. Although this was manifestly of human *level*, any pretensions that it might have to chess intelligence of human *type* would need to be separately assessed, according to its ability to comment intelligently on its own and other games. The world still waits for the first chess machine to make a credible showing on this score.

Towards the strong form: the Turing-Newman Test

In the largely unknown closing section of the 1950 *Mind* paper, entitled "Learning Machines", Turing turns to issues more fundamental than intuitional semantics:

> We may hope that machines will eventually compete with men in all purely intellectual fields. But which are the best ones to start with? It can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. This process could follow the normal teaching of a child. Things could be pointed out and named, etc.

What time-scale did he have in mind for his "child-machine project"? Certainly not the 50-year estimate for his game for disconcerting philosophers. He and Max Newman consider the question in a 1952 radio debate (Copeland, 1999):

> Newman: I should like to be there when your match between a man and a machine takes place, and perhaps to try my hand at making up some of the questions. But that will be a long time from now, if the machine is to stand any chance with no questions barred?
>
> Turing: Oh yes, at least 100 years, I should say.

So we are now half-way along this 100-year track. How do we stand today? The child-machine prescription segments the task as follows:

Stage 1. ACCUMULATE a diversity of generic knowledge-acquisition tools, especially perhaps those designed to exploit already accessible large knowledge sources such as the Web.

Stage 2. INTEGRATE these to constitute a "person" with sufficient language-understanding to be educable, both by example and by precept.

Stage 3. EDUCATE the said "person" incrementally over a broad range of topics.

Step 1 does not look in too bad shape. An impressive stock-pile of every kind of reasoning and learning tool has been amassed. In narrowly specific fields, the child-machine trick of "teaching by showing" has even been used for machine acquisition of complex concepts. Chess end-game theory (*see* Michie 1986, 1995) has been developed far beyond pre-existing limits of human understanding. More recently, challenging areas of molecular chemistry have been tackled by the method of Inductive Logic Programming (e.g. Muggleton, S.H., Bryant, C.H. and Srinivasan, A., 2000). Again, machine learners here elaborated their insights beyond those of expert "tutors". Above-human intelligence can in these cases be claimed, not only in respect of performance but also in respect of articulacy (*see* Michie, 1986).

A feeling, however, lingers that something crucial is still lacking. It is partially expressed in the current *AI Magazine* by John Laird and Michael van Lent (2001):

> Over the last 30 years, research in AI has fragmented into more and more specialized fields, working on more and more specialized problems, using more and more specialized algorithms.

These authors continue with a telling point. The long string of successes, they suggest, "have made it easy for us to ignore our failure to make significant progress in building human-level AI systems". They go on to propose computer games as a forcing framework, with emphasis on "research on the AI characters that are part of the game".

To complete their half truth would entail reference to real and continuing progress in developing *generic* algorithms for solving *generic* problems, as in deductive and inductive reasoning; rote learning; parameter and concept learning; relational and object-oriented data management; associative and semantic retrieval; abstract treatment of statistical, logical, and grammatical description and of associated complexity issues, and much else. None the less, the thought persists that something is missing. After all we are now in 2001. Where is HAL? Where is even proto-HAL? Worse than that: if we had the HAL of Stanley Kubrick's movie "2001, a Space Odyssey" would we have reached the goal described by Turing, a machine able to "compete with men in all purely intellectual fields"?

Seemingly so. But is that the whole of the goal we want? Should an associated goal not be added, namely to " *co-operate* with men (and women) in all purely intellectual fields", as intelligent humans also do with no more than a few known pathological exceptions? Impressive as was the flawless logic of HAL's style of reasoning in the movie, the thought of having to co-operate, let alone bargain, let alone relax, with so awesomely motivated a creature must give one pause. The picture has been filled in by John McCarthy's devastating satirical piece "The Robot and the Baby", available through

`http://www-formal.stanford.edu/jmc/robotandbaby.html`

The father of the logicist school of AI here extrapolates to a future dysfunctional society some imagined consequences of pan-logicism. By this term I mean the use of predicate logic to model intelligent thought unsupported by those other mechanisms and modalities of learning and reactive choice which McCarthy took the trouble to list in his 1959 classic "Programs with common sense" (see also Michie, 1994, 1995).

The hero of McCarthy's new and savage tale is a robot that, from an impeccable axiomatization of situations, actions and causal laws, applies mindless inferences to an interactive world of people, institutions and feelings. The latter, moreover, are awash with media frenzy, cultism, addiction, greed and populism. Outcomes in McCarthy's story, are at best moderate. How should an artificial intelligence be designed to fare better?

Implicitly at least, Stage 2 above says it all. Required: a way to integrate the accumulated tools and techniques so as to constitute a *virtual person*, with which (with whom) a user can comfortably interact, "a 'person' with sufficient language-understanding to be educable, both by example and by precept". Running a little ahead of our theme, the reader will find that initial experimentation with Step 2 has taken us to ground-level educability by precept, but not yet to educability by example.

Logic then places on the shoulders of AI a new responsibility, unexpected and possibly unwelcome: we have to study the anatomy and dynamics of the human activity known as chat. Otherwise attempts to simulate the seriously information-bearing components will fail to satisfy needs extending beyond information-exchange to what is known as rapport.

The ability to converse is of course, in itself not a major form of intelligence. In the same way it would insult an aerobatic performer to suggest that maintenance of level flight is a major form of his art. And yet a level-flight capability is necessary. First, level flight has to figure early rather than late in the making of a stunt flyer. Second, in executing their repertoires, exhibition flyers must repeatedly pass through level flight, if not as common link between one exotic manoeuvre and the next, then at least as beginning and end states of the exhibition as a whole. Disjointed sequences of spectacular trick-flying are not aerodynamically viable. For 45 years, mainstream AI has gambled on the hunch that the same does *not* apply to the aerobatics of the intellect. With the commercial rebirth of his 1950 imitation game, Turing's opposite prescription is surely now due for a look, namely his still-to-be-attempted "child machine" approach.

If a person cares to reject that prescription, then a socially chilling agent such as HAL, built according to some different implementational strategy, could conceivably succeed. That is, a chat-incapable but linguistically adequate HAL might on half the occasions be mistaken for the human in imitation-game contests. Although to some the endeavour may seem unlikely, it would deserve encouragement and support.

However, if the child-machine prescription is also to be tried, an infant Hal devoid of chat propensity will not do. School and university teachers, and human coaches in other areas of skill can be hired to devote their talents to building the budding HAL's knowledge of the world. But then they are in effect being asked to tutor the software equivalent of a brainy but autistic child. Empirical findings have accumulated from protracted and repeated attempts to teach such children. Their otherwise excellent mental models of situations, actions and causal laws give them predictability of objects but not of the movements, motivations or feelings of humans. Hence they lack dependable channels of communication with would-be teachers. It seems as close to inescapable as anything could be that a software agent devoid of social and co-operative intelligence could not constitute a feasible stepping stone from Stage 2 to Stage 3.

Rapport maintenance

To see how to do Step 2 (integration into a user-perceived "person") is not straightforward. Moreover, what we expect in a flesh-and-blood conversational agent comes more readily from the toolkit of novelists than of computer professionals.

1. Real chat utterances are mostly unparseable. They are concerned with associative exchange of mental images. They respond to contextual relevance rather than to logical or linguistic links [1].

2. A human agent has a place of birth, age, sex, nationality, job, hobbies, family, friends, partners; plus a personal autobiographical history, recollected as emotionally charged episodes; plus a complex of likes, dislikes, pet theories and attitudes, stock arguments, jokes and funny stories, interlaced with prejudices, superstitions, hopes, fears, ambitions etc [2].

3. On meeting again with the same conversational partner, a human agent recalls the gist of what has been divulged by both sides on past occasions [3].

4. Crucially for implementers, a human agent typically has *goals* beyond mere chat, whether fact-transmission, or fact-elicitation, or persuasion, or other.

5. A human agent remains ever-ready to default to chat-mode to sustain rapport [4].

Background findings in cognitive neuroscience generally are surveyed in Ramachandran and Blakeslee's (1999) highly readable paperback.

Some researchers may make the point that "real rapport" (which might be required to pass the strong version of the imitation game) will be a good deal harder to achieve than illusory or surface rapport. But the rapport here aimed at is a sensation or mood induced *in the user*, the reality of which only he or she can judge. If the user announces "With Sophie, I feel *real* rapport!" it is hard to know where next to take the argument.

Recent experimentation

Over the last two years Claude Sammut and I have begun experimentally to develop and test activation networks of pattern-fired rules, hierarchically organized into "contexts". Our first deliveries and continuing enhancements have been to the Sydney Powerhouse Museum's permanent exhibition "Cyberworld". When run in text-only mode, our product uses "Sophie" as her stage name. But for the Powerhouse interactive exhibit the Californian company Virtual Personalities Inc. generously contributed under license their "Sylvie" face-and-voice animation. We interfaced this to a copy of the Infochat engine together with a suite of scripts to constitute an agent able to move seamlessly back and forth between "goal mode" (conveying background on the "Cyberworld" exhibits) and "chat mode".

Before sampling the public's conversations with the exhibit, my psychology colleague Dr Hayes Michie ran pilot useability tests using student judges. Student volunteers were recruited from the UNSW Psychology Department. Each had two 7-minute chat sessions, text-only, punctuated with questionnaires about their experience. To rehearse procedures, tests of Loebner chat agents were all run with research assistant Stephanie Compton as subject. The results of these were treated as only anecdotally informative. The tabulation reproduced below below is the questionnaire summary sheet referring to Sophie's pilot results.

In the post-pilot phase all subjects will be tested under similar conditions both with Sophie and with each of a small battery of chatbots judged to approach Sophie in performance. These will include recent winners and runners-up in the annual Loebner Prize competition, together with one or two of the commercial market leaders. In addition the same panel of subjects will be asked to score the two or three longest of the month's accumulated transcripts of the Museum chatbot guide's interactions with members of the public.

SOPHIE: Steph, it's my pleasure! Before I get into detailed things,... Cyberworld is about computers and communications. That includes robots, telecommunications and other stuff. I specialize in the material on Alan Turing, Turing machines, and his ideas on Artificial Intelligence and Learning. Anything appeal? There's more ...

STEPH: Like what?

SOPHIE: Yes, well Steph before you commit, there's a great exhibit on Charles Babbage. There's Simulation, and added to that... I forgot some other things, multimedia for instance. And process control, which I didn't mention before. So what's the final decision?

STEPH: Well, what about Charles Babbage?

SOPHIE: In the mid 1800's, the English mathematician, Charles Babbage, designed a general purpose mechanical calculating machine that anticipated the principles and structure of the modern computer. Do you know why he was never able to finish building his Analytical Engine and why we didn't have mechanical computers 100 years earlier?

STEPH: Because he did not have the resources available that we have?

SOPHIE: In that ballpark. The only technology Babbage had available to him consisted of wheels, gears and steam power. His calculating engines were more complicated than anything anyone had yet built.

STEPH: Wow, he must have been pretty smart.

SOPHIE: OK, tell me. Why?

STEPH: Because he had all these brilliant ideas and visions. His mind must have been very advanced for his time.

...

Stephanie reported the experience as OK but "boring". Her later experience, after the chat links were in place, struck her quite differently. She burst into excitement about her new companion. The following is excerpted form the first of the new sessions. Note the transitions in and out of chat mode.

SOPHIE: What next? My menu says AI, Turing, Stumpy,Universal Machine, Media, Babbage, Simulation, Control, Interfaces,... I think that's it. Or we could just chat a while.

STEPH: Why don't we chat?

SOPHIE: That's nice that you take an interest. Do go on. Do you have any brothers and sisters?

STEPH: I have three brothers and three sisters. What about you?

SOPHIE: Great. So shall we just have a gossip, Steph?

STEPH: Yes

SOPHIE: Why don't you kick off, Steph? Tell me something about your views and interests.

STEPH: Well, I like to play sport, listen to music, go to the beach and see my boyfriend.

SOPHIE: Yeah, but stewing indoors can be quite fun, – within the family circle, though. How is it in your family? Any others share your tastes?

STEPH: Well, most of my family like going to the beach and camping so we all get on really well.

SOPHIE: I grew up in Pennsylvania on the farm which my parents still run, these days helped by uncle Albert. Shall I tell you, then, how sister Julia and I ended up in Australia?

{. . .

[LATER SOPHIE TERMINATES THE "CHAT" INTERLUDE]

. . .

SOPHIE: You have a friend in every port, it seems to me.

STEPH: Well, kind of.

SOPHIE: We'll come to that. Meanwhile how about some exhibition topics. That's what you came for surely?

STEPH: Yeah, I suppose we better get back to that.

. . .

The above excerpt was generated by the same versions of engine and scripts as those installed for public use in summer 2001 at Sydney's Powerhouse Museum, and those used in the tabulated usability tests.

How does the system connect back and forth between serious dialogue and chat? Switch to a chat file is triggered by any pattern in the user input which indicates such a wish: in the foregoing example quite simply by Stephanie's "Why don't we chat?" Such switching has to be strictly contingent on user input patterns. No-one wants the agent to start chatting at a moment of goal-directed significance, e.g. at the "take it or leave it" point in a negotiation.

Triggers for getting back from chat mode to serious mode are many and various. In the above example the trigger was Sophie's failure to find a match for Stephanie's "Well, kind of." Such potentially embarrassing eventualities can usually be covered by prevarication followed immediately by concern about the neglected instructional goal. The user can then dismiss, or be distracted from, the impression that the agent was not paying proper attention, and focus instead on whether or not she wants to return to the main goal. It was open to Stephanie to reply with something like "Maybe we should, Sophie. But talk about boyfriends is more fun. Do you ever think about marriage?"

Pattern-matches will here be picked up by "maybe", "boyfriend" and "you * marr* " (* stands for wild-card). "Maybe" belongs to a macro of patterns standing for only equivocal affirmation. So either "boyfriend" or "you * marr* " will win. As their author, I know from the scripts that if the first scores higher and Sophie hasn't yet mentioned boyfriends, then she will open the subject of her Indian boyfriend Max who studies physics at UNSW and has emotional difficulties. If however the "you * marr* " pattern wins, then she will produce some variant of:

> One day, perhaps ... I've read that unhappy marriages come from husbands having brains. Max is fearfully brainy.

Stephanie is almost certain now to ask something about the newly-mentioned "Max". So a new phase of chat is launched.

The mechanism driving the chat dialogue of NativeMinds' customer service facility (*see* Appendix 1) is of an altogether less sophisticated order. It is appended simply to exhibit performance from one of the world's leading commercial players. Note that the company CEO's own claims on this score are modest: "Our vRep has to be able to answer 200 questions that we call banter. They're about the weather, are you a boy or girl, will you go out with me? They're ice breakers." In the Sophie project we envisage an eventual need to field a system of many thousand chat-response rules, additional to those developed for application-specific information exchange.

The current prototype has a few hundred. Note incidentally that where application goals include elicitation of personal profiles of users, as in market research, the distinction between the two modes becomes blurred.

Management of achievement goals

As is evident, the program's default mode is to use S-R behaviour, where user utterances that match certain kinds of pattern constitute one kind of stimulus. It also has a "goal" or "direction" construct that provides coherence. The latter is effected through liberal use of global variables that may be set, overwritten or unset by the firing of any rule. Collectively they define the agent's fixed and current states of belief, not only of objective fact (knowledge) but about attitudes and feelings. The latter can include not only moods attributed to self or user, but also what she knows about the user's own fixed and current beliefs, for example the user's birth-place (fixed), his or her favourite sport (fixed) or his or her prediction of an election result (current). From the coherence point of view, the most important categories are concerned with the following:

§ Who am I talking to and what did I earlier know about him or her?

§ What new facts have I gleaned since this particular session started?

§ What facts new to this user have I imparted during the same period?

§ From which topic file (context) did we arrive?

§ What topic file (context) are we currently in?

§ In the current file, have we yet reached stage 1, stage 2, stage 3?

Coupled with all of this is a key feature of rule structure which permits a rule to fire if and only if a specified boolean condition is satisfied over a subset of the agent's current knowledge. Use of this feature for goal management rests on a scripting discipline that associates with each topic file a logical profile of things that must be known for the agent to become ready either to switch back to the topic file (context) from which she came, or definitively to quit the current file for a new conversational departure. For instance, the "getname" topic file that forms part of the initial greeting sequence will not return control to the greeting file until either a user-name variable or a no-name variable has been set. By the same token, rules of the greeting file will never switch control to getname if user-name is already set. At end of session, all global variables are saved. At the start of a new conversation, once the agent has identified the user as someone previously known to her, the corresponding set of variables are automatically read in.

Chat and ballroom dancing

In the quoted fragments we glimpsed an alternation between front-stage business and back-stage chat. The latter is a social activity analogous to "grooming" in other primates. The surprise has been the indispensability of grooming, - the really hard part to implement. Yet this result might have been inferred from the extensive studies of human-machine interactions by Reeves and Nass (1996).

In an important respect, human chat resembles ballroom dancing: one sufficiently wrong move and rapport is gone. On the other hand, so long as no context violation occurs, "sufficiently" turns out to be permissive. In the above, Sophie ignores a direct question about her brothers and sisters, but stays in context. If not too frequent, such evasions or omissions pass unnoticed. When the human user *does* pick them up and repeats the question, then a straightforward reply is essential, even if it amounts to no more than an apologetic admission of ignorance. In some matters, such as family details, ignorance is not credible. In such a case urgent script repair is indicated.

Capable scripting tools, such as those that Sammut has largely been responsible for pioneering, make incremental growth of applications a straightforward if still tedious task for scripters. Adding and linking new rules to existing topic files, and throwing new topic files into the mix can proceed without limit. Addition of simple SQL database facilities to our *Pattern-Script* language has been proved in the laboratory and is currently awaiting field testing. Agent abilities to acquire and dispense new facts from what is picked up in conversation will thereby be much enhanced.

Since then a new wing of the project led by Dr Zuhair Bandar of the Department of Computing and Mathematics, Manchester Metropolitan University (MMU), has begun further enhancement of our Infochat™ scripting technology and documentation. A commercial company Convagent Ltd has been set up in the UK. Stockholders include the MMU research workers, the University itself, the Human-Computer Learning Foundation, and Dr Akeel Attar, Britain's leading practitioner of commercial machine-learning software for knowledge engineering.

Forward look

Incorporation of serious machine learning facilities in the Sophie agent remains for the future. At present the agent can learn facts from conversations, and can accumulate the rote knowledge indefinitely over successive sessions. Rough and ready structuring of globally stored facts is automatically effected by context-restriction mechanisms. Certain obvious opportunities to embed re-inforcement learning by the agent are ready for future implementation at such time as human resources permit. Such learning concerns the incremental favouring of empirically more effective over less effective ways of constructing responses. The operational criterion to be used for “more effective” is “provoking on average higher frequencies of easily pattern-matched responses”. Luxury features, however desirable, such as concept-learning, and logical as opposed to associative inference, may have to wait a year or two before receiving serious consideration.

The second of this series of papers will include the main features of the scripting engine, including a formal specification of version 1 of the PatternScript language. The present early prototype is informally but comprehensively documented in Michie and Sammut (2001). Our return-on-investment principle of project scheduling currently leads us to concentrate on bulk scripting, on scripting methodology and its experimental validation, and in particular on the refinement and extension of principles of scripting in the light of the stream of new data. Our installed agent at the Sydney Powerhouse Museum now generates dozens of logged visitor-agent conversations per day. With all this goes the development of support software, including the development of graphical editors and profilers, and automatic capture and statistical summary of chat sessions.

A word, finally, on ascribing mental qualities to machines. Clearly people *do* habitually do this. Whether and in what circumstances such ascription is necessary for useful human-machine interaction is a question that the commercial rise of chatbots has passed to the market-place. For the scripter, the human propensity is simply a given, to be empirically studied, - as in the earlier tabulated pilot test conducted by Dr Hayes Michie.

The importance of characterizing the propensity emerged when we were able to compare the pilot sample of some ten conversations and the first crop harvested from the Powerhouse site. The contrasts were stark.

First, Hayes Michie’s subjects, who were all Psychology undergraduate students, had no difficulty with Sophie’s opening utterance, which was “Ready?” Some answered affirmatively, others asked for more time, but all responded. Yet the majority of the Powerhouse users immediately terminated the interaction. Simple variation of the script should yield clues as to which, if any, of the following contrasts forms the leading term:

Education: Powerhouse users were mainly children and school leavers. Pilot users were first-year University students.

Posture: Powerhouse users were ambulant. Pilot users were seated.

Mood: Powerhouse users were on a leisure trip. Pilot users were doing a (volunteer) job.

Second, in the pilot sample there was no obvious difference attributable to the user’s gender. But there was a large gender difference in the non-quitting Powerhouse sample (after “Ready?” comes an enquiry for the person’s name, usually sufficient to disclose gender). Female users were mainly chatty. But a substantial proportion of the males were persistently abusive and obscene. If the differential is confirmed and continues to be large, scripts will, if this proves possible, be customized to age and gender with the aim of for calming down male teenagers faced with animated virtual females.

Although these are early days the market already holds the key to developments. But there is something more important even than the market. Success in the difficult task of chat-simulation is preconditional to a future in which computers gain information and understanding *through interaction* *with lay users.* If this means that computer professionals will one day cease to be the sole and necessary channels for information technology's further enrichment, then so be it.

Acknowledgements

My thanks are due to the Universities of Edinburgh and of New South Wales for funds and facilities in support of this work, and also to my fellow Trustees of the HCL Foundation for endorsing contributions by the Foundation to some of the work's costs. I also owe a particular debt to my friend Nils Nilsson for thorough and illuminating critiquing of an earlier draft of the paper, and an overwhelming obligation to my coworker Claude Sammut, who largely designed and built the C-code Infobot kernel used to develop the present prototype. The developers' interface that chiefly differentiates the Infochat product from this original kernel was primarily the programming work of Robert Rae, of the AI Applications Institute, University of Edinburgh, UK. I also gratefully acknowledge further work on developers' support software by Jamie Westendorp of the AI Lab of the University of New South Wales, Australia, and by David McLean of the Manchester Metropolitan University, Manchester, UK.

By reason of my personal representation on the Board of Convagent Ltd of the Foundation's commercial interest, it is proper that I should also declare it here.

References

Bratko, I. and Michie, D. (1980) An advice program for a complex chess programming task. *Computer Journal,* **23** (4), 353-359.

Copeland, B.J. (1999) A lecture and two broadcasts on machine intelligence by Alan Turing. In *Machine Intelligence 15* (eds. K. Furukawa, D. Michie and S. Muggleton), Oxford: Oxford University Press.

Dennett, D. (2001) Personal communication.

Huberman, B. (1968) A program to play chess end games. *Technical Report* no. CS 106, Stanford University: Computer Science Department.

Laird, J.E. and van Lent, M. (2001) Human-level AI's killer application: interactive computer games. *AI Magazine,* **22** (2), 15-25.

McCarthy (1959) Programs with common sense. In *Mechanization of Thought Processes,* Vol. 1. London: Her Majesty's Stationery Office. Reprinted with an added section on situations, actions and causal laws in *Semantic Information Processing* (ed. M. Minsky). Cambridge, MA: MIT Press, 1963.

Michie, D. (1986) The superarticulacy phenomenon in the context of software manufacture. *Proc. Roy. Soc. A,* **405,** 185-212. Reprinted *in The Foundations of Artificial Intelligence: a source book* (eds D. Partridge and Y. Wilks), Cambridge: Cambridge University Press.

Michie, D. (1994) Consciousness as an engineering issue, Part 1. *J. Consc. Studies,* **1** (2), 182-95.

Michie, D. (1995) Consciousness as an engineering issue, Part 2. *J. Consc. Studies,* **2** (1), 52-66.

Michie, D. and Sammut, C. (2001) *Infochat Scripter's Manual,* Manchester, UK: Convagent Ltd.

Muggleton, S.H., Bryant, C.H. and Srinivasan, A. (2000) Learning Chomsky-like grammars for biological sequence families, *Proc. 17th Internat. Conf. on Machine Learning,* Stanford Univ. June 30th.

Ramachandran, V.S. and Sandra Blakeslee (1998) *Phantoms in the Brain: Human Nature and the Architecture of the Mind.* London: Fourth Estate (republished in paperback, 1999).

Reeves, B. Nass, C.I. (1996) *The Media Equation: how people treat computers, televisions, and new media like real people and places.* Stanford, CA.: Center for the Study of Language and Information.

Sammut, C. (2001) Managing context in a conversational agent. *Machine Intelligence 18*, this volume.

Turing, A.M. (1950) Computing machinery and intelligence. *Mind,* **59** (236), 433-460.

Turing, A.M. (1952) in *Can Automatic Calculating Machines be Said to Think?* Transcript of a broadcast by Braithwaite, R., Jefferson, G., Newman, M.H.A. and Turing, A.M. on the BBC Third Programme, reproduced in Copeland (1999).

Appendix: New Customer Service Software

By Sabra Chartrand

Walter Tackett, chief executive of a San Francisco company called NativeMinds, has patented a software technology called automated virtual representatives, or vReps. It conducts customer service, sales and marketing for online businesses.

The vReps are computer-generated images - sometimes animation, sometimes photos of real models - that answer customer questions in real time using natural language. Users type in their questions, and the responses appear on screen next to the image of the vReps. Mr. Tackett and the co-inventor, Scott Benson, say the technology can mimic and even replace human customer service operators at a fraction of the cost, whether a business has traditionally used phone, fax, e-mail or live conversation to deal with customers.

The invention came about from research NativeMinds conducted on what consumers and companies wanted from a virtual customer services force. Consumers, the company found, did not care whether the character contained enormous amounts of universal knowledge; they just wanted fast, accurate answers in their specific subject area. Companies wanted virtual customer support software they could easily maintain. They did not want to hire a computer engineer to run the program.

“They want to be able to put a code monkey on it,” Mr. Tackett explained. “That's a liberal arts major involved in HTML or Java, someone not formally trained in computer science or as an artificial intelligence or natural language expert.”

So Mr. Tackett and Mr. Benson developed software based on pattern recognition and learning by example.

"The key thing is to get the user not to pick up the phone and talk to a person," Mr. Tackett said. "The key to that is to get the vRep to answer all the questions that can be reasonably answered and have a high probability of being correct."

To do that, the patented software starts with the answers and works backward. A vRep might be programmed with thousands of answers, each of which has a set of questions that could prompt the answers. Each answer could have dozens of questions associated with it. The number depends on how many ways the query could be phrased.

"The examples are archetypes or prototypes of inputs that should trigger an answer," Mr. Tackett said. "The invention runs each example through the system as if someone has put it in. The paradigm we typically use is learning by example. Here's what we want the vRep to say, and we give an example of how people may phrase their question to get that output." For example, someone might ask, 'Who the heck is this Walter guy?' Or, 'Tell me about Walter,'" he said, referring to himself. The system comes with a self-diagnostic, he added, so that it can "take all the examples it ever learned and verify that it still remembers them correctly."

The self-test is to prevent information from one area generating an incorrect answer in another. "Someone might ask, 'Who is the president?'" he said. "That could be a question no one has ever asked before. They might mean, 'Who is the president of the U.S.?' But the system would say, 'Walter.' This is a classic problem of vReps." The self- test would periodically find and eliminate incorrect answers, based on the responses that users provide, he said.

Companies like Coca-Cola, Ford and Oracle are using the vReps software for various functions on their Web sites. Mr. Tackett said research had determined that virtual representatives could save money, an aspect that surely appeals to embattled e-businesses.

"A vRep costs less than a dollar a conversation, while Forrester Research has pegged phone calls to a real customer service person at an average of $30 each," Mr. Tackett said. "With a vRep, the length of the conversation doesn't affect the cost because it's maintained by one person," he added.

Not all of the programming is technical or product-oriented. "Our vRep has to be able to answer 200 questions that we call banter," Mr. Tackett said. "They're about the weather, are you a boy or girl, will you go out with me? They're ice breakers."

He and Mr. Benson received patent 6,259,969.

Footnotes

[1] It is of interest that congenital absence of the capacity to handle grammar, known in neuropsychology as "agrammatism", does not prevent the sufferer from passing in ordinary society. Cases are ordinarily diagnosed from hospital tests administered to patients admitted for other reasons. Ultimately, of course, chat agents will have to extract whatever can be got from linguistic analysis. It is a matter of priorities.

[2] Disruption of this cohesive unity of personal style is an early sign of "Pick's disease", recently linked by Bruce Miller and co-workers with mal-

function of an area of the right fronto-temporal cortex. Reporting to a meeting in early summer 2001 of the American Academy of Neurology meeting in Philadelphia Miller presented results on 72 patients. One of them, a 63-year-old woman, was described as a well-dressed life-long conservative. She became an animal-rights activist who hated conservatives, dressed in T-shirts and baggy pants and liked to say "Republicans should be taken off the Earth!"

[3] Failure of this function in humans is commonly associated with damage involving the left hippocampal cortical area.

[4] Reverting to the child-machine concept, how much of value and use could a school teacher impart to a child with whom rapport was impossible? In clinical practice a condition of this kind is found in "autism". Children with the disorder are indeed found to be almost unteachable.