Linköping Electronic Articles in
Computer and Information Science
Vol. 5(2000): nr 30

It takes two (at least) to tango

Donald Michie

Linköping University Electronic Press
Linköping, Sweden

http://www.ep.liu.se/ea/cis/2000/030/

Published on December 21, 2000 by
Linköping University Electronic Press
581 83 Linköping, Sweden

Linköping Electronic Articles in
Computer and Information Science
ISSN 1401-9841
Series editor: Erik Sandewall

Typeset by the author using LaTeX
Formatted using étendu *style*

Recommended citation:
<Author>. <Title>. Linköping Electronic Articles in Computer and Information Science, Vol. 5(2000): nr 30. http://www.ep.liu.se/ea/cis/2000/030/. *December 21, 2000.*

This URL will also contain a link to the author's home page.

Discovery as Collaboration

In his *Structure of Scientific Revolutions*, Thomas Kuhn distinguishes between:

(1) "discoveries, or novelties of fact" and

(2) "inventions, or novelties of theory".

Examples that have been cited of the first are the discovery of the New World by Columbus, the discovery of the planet Neptune, the discovery of new chemical elements and the discovery of penicillin.

Examples of the second category are Newton's theory of gravitation, Darwin's theory of natural selection, the microbial theory of infectious disease and the greenhouse theory of global warming.

Traditional thought holds that theory discovery typically proceeds in discontinuous "insights of reason" (Aristotle), "creative intuitions" (Bergson), "inspirations" (Popper), "flashes of genius" (Reichenbach) and the like. These "interventions from nowhere" lie beyond immediate introspection. They are commonly ascribed to intrinsically mysterious and possibly unknowable causes.

A similar position *could be* (but is not) taken about fact discovery. Householders who have forgotten where they have left something not uncommonly testify that their wanderings took them "for no particular reason" to the approximate whereabouts of the lost object. When discovery immediately follows, protagonist and onlookers do not ordinarily refer to a "hidden hand" or to a "divine spark". Similar restraint moderates the pride of fact-discoverers of greater moment. The following is early Portuguese testimony to the theme of background knowledge to which we shall later return:

> *They discovered new islands, new lands, new seas, new peoples, and what is more important, new heavens and new stars. Now it is clear that these discoveries were not achieved through guesswork: our seamen set off well trained and provided with instruments and rules of astronomy and geometry.*
>
> *from Pedro Nunes, 1537.*

Person and Zombie

Household fact-discovery is sufficiently mundane that there is no threat to human dignity in supposing that perceptions and inferences come from below the level of the conscious Person. What neuroscientists tend to refer to as "the Zombie" probably guided the Person's steps, as the only means of signalling. But in the grander arena of theory-discovery, the mind's sudden insights are more often presented as coming from nowhere.

The thrust of this paper is that software simulation of theory-discovery, just as much as of fact-discovery, depends on abandoning the model of a unitary mind visited by flashes. Modern brain science views the overwhelming bulk of human perceptions and behaviours, including those of abstract thought, as driven in fast time by hidden zombies. The Person reserves awareness for a few strategically critical functions of planning, monitoring, rationalization and communication.

The singular noun "Zombie", like "Army", here denotes myriads of unseen agents, mutually co-ordinated. As in Tolstoy's portrayal of Napoleon's role in battle, the Person's reward for being swept along by the stream of events is the privilege of issuing commands and requests for information, and of exchanging messages with allied generals (other Persons). Tolstoy's contemporary, the British scientist Francis Galton (1883), was already describing the conscious mind in similar terms:

> *The position of consciousness appears to be that of a helpless spectator of but a minute fraction of automatic brainwork.*

In the time-scales of professional tennis or lightning chess Galton's account is today known not to be an exaggeration. In slower time, though, the Person's role is a good deal more than that of mere commentator. Proactive collaboration by the Person with the hidden zombies then emerges, and low-capacity channels between the two open up. The following specimens of signalling between the two are consistent with modern knowledge.

From Person to Zombie (used plurally, as in Army)

- Direct invocation: "What's 17 minus 3?"
- Setting goals: "Let's open the fridge."
- Setting constraints: "Stooped posture for faster bowing of cello."

From Zombie to Person

- Direct answer to fact-query: Person "sees in a flash" that 17 minus 3 just *is* 14.
- Attracting Person's attention through internal body signals: "Hungry!"
- Person-observable signals broadcast by automatized movements: "Bowing still too slow."
- Other (e.g. "blind-sight").

The cello-playing example is from K. Furukawa's discussion elsewhere in this Volume.

For *cognitive simulation,* opacities found in nature must be preserved. But from the Person's point of view, the Zombie's opacities can be inconvenient. For the technological purposes of *machine intelligence* we can aim higher. Experimental demonstration of the extraction of transparent "behavioural profiles" from a fast real-time skill was shown by Michie, Bain and Hayes-Michie (1990; see also Michie, 1998). In effect, it is possible to populate even the lowest levels of fast-time problem-solvers with "super-articulate zombies." A paradigmatic case study from A. D. Shapiro (1987) will later be presented.

The phenomenon of "blind-sight" will now be described.

A case of "blindsight"

The following clinical condition is illustrative of Person-Zombie duality. When the primary visual cortex has been compromised by stroke, trauma or surgery, the patient is partially or wholly blind. But the phylogenetically more primitive "orienting" visual pathway is still intact. Even though the Person is unaware of what is going on, this pathway can capture and utilize visual inputs, thereby generating tell-tale behaviours in the seemingly blind. Here is a patient of Sanders and Weiskrantz, referred to in Ramachandran's (1998) *Phantoms in the Brain* as Drew.

They held up a stick, in either a vertical or a horizontal position, in his blind field and asked him to guess which way the stick was oriented. Drew had no problem with this task, although he said that he could not see the stick. After one such long series of "guesses", when he made virtually no errors, he was asked, "Do you know how well you have done?"

"No," he replied, "I didn't - because I couldn't see anything; I couldn't see a darn thing."

Note that the problem posed by the experimenter's guessing game was necessarily solved by a collaboration between two conceptually and experimentally separable agents of the patient's brain. The Zombie made the stick-orientation discoveries. Drew (the Person) somehow unknowingly glimpsed them, before retransmitting them as guesses.

Duality Principle

Muggleton and Michie (1996) have proposed a somewhat similar two-layer software architecture for AI, referring to it as "The Duality Principle":

... software involved in human/computer interaction should contain two distinct layers a declarative knowledge-level layer and a lower-level functional or procedural-knowledge layer. This extends the formal methods separation of specification and implementation by requiring that the declarative layer be capable of extensive human interrogation at run time.

An advantage of the artificial over the natural system is that the knowledge-level Person can be endowed with better communication with run-time zombies than is enjoyed by the conscious brain. In the case of the forgetful householder, a common pattern is for the person to review in declarative memory what can be recalled of the circumstances surrounding the last sighting of the lost object. Having found it, with or without subconscious promptings, in the mental model ("Aha!"), the person then launches appropriate zombies to support real-time exploration of the newly recalled vicinity. In intelligent software, the two-layer operating systems of the future will doubtless short-circuit such mechanisms. Direct inspection of traces not only of top-level transactions but (critically) of histories and states of subordinate agent-modules could become routine.

One of our agents is missing

The foregoing stick-guessing example illustrates the brain's duality principle from the perceptual side. Below is an action-oriented case chosen to illustrate a form of fact discovery usually known as problem-solving. Again it comes from Ramachandran (1998). The problem is simple arithmetic. The patient had suffered a small localized stroke affecting the brain's left angular gyrus, already known to be specifically associated with numerical calculation.

At one point I said, "Okay, Bill, can you subtract seven from one hundred? What's one hundred minus seven?"

He said, "Oh, one hundred minus seven?"

"Yeah."

"Hmmm, one hundred minus seven."

"Yes, one hundred minus seven."

"So," said Bill. "One hundred. You want me to take away seven from one hundred. One hundred minus seven."

"Yes."

"Ninety-six?"

"No."

"Oh," he said.

"Let's try something else. What's seventeen minus three?"

"Seventeen minus three? You know, I am not very good at this kind of thing," said Bill.

"Bill," I said, "is the answer going to be a smaller number or a bigger number?"

"Oh, a smaller number," he said, showing that he knew what subtraction is.

"Okay, so what's seventeen minus three?"

"Is it twelve?" he said at last.

Further questioning revealed that Bill had an intact and quite sophisticated model of the number system, including infinity, approximation, relative magnitudes and the rest. Only the calculational agent was missing from the library of procedures, leaving the library manager (Bill) without resource whenever this specific function was tested.

Imagine, then, a Turing Test confined to the domain of simple arithmetic. The respective stances of the logicist school ("let an artificial Bill work it out from first principles") and of the reactivist school ("leave everything to the zombies") both fail.

- For non-trivial problems the logicist solution fails the primary task, as did Bill.
- The reactivist solution makes a no-show in the associated discussion, at which Bill shone.

Child Machine

Now unite Person and Zombie in a partnership, with one additional responsibility for the Person, namely *zombie-training*. Result: a three-layer prescription for skill-acquisition reminiscent of Turing's (1950) "child machine". Turing asked why we should not break the task of developing a machine intelligence into two steps, essentially:

1. Build a teachable machine.
2. Teach it.

A modern up-date might further refine this:

0. Build a teachable machine (a task for human developers).
1. Educate its declarative layer (a task for human teachers).
2. Train its functional layer (a task for human trainers)

Layer 1 must incrementally take over some of the human teacher's tasks, acquiring further facts and generalizing theories for itself. Layer 1 must also take a hand in the building of layer 2, i.e. in zombie-training. By this we mean that it must acquire certain declarative elements of the human trainer's craft concerned with setting the framework of task-segmentation, subgoals and constraints within which level-2 re-inforcement learning and classification learning proceeds.

Returning to Bill, long before his calculating zombie was disabled by the stroke, he had early lost conscious touch with it - as we all do once arithmetic skills have become automatic. Yet his youthful Person, goaded and guided by teachers, had personally trained this zombie in the first place, by setting it goals, rehearsing it in endless practice, and using declarative-layer knowledge to correct errors.

News from nowhere

So what of novelties of theory rather than novelties of mere fact? Testimony of creative thinkers, from the novelist Henry James to the mathematician Henri Poincare is in line with the idea that flashes of insight (illuminating the declarative layer) are preceded by systematic and repetitive thought-experiments accomplished by largely unconscious brainwork in the functional layer:

> *For a fortnight, I struggled to prove that no function analogous to those I have since called Fuchsian could exist; I was then very ignorant. Every day I sat down at my work table where I spent an hour or two; I tried a great number of combinations and arrived at no result. One evening, contrary to my custom, I took black coffee; I could not go to sleep, ideas swarmed in clouds; I sensed them clashing until, as it were, a pair would hook together to form a stable combination. By morning I had established the existence of a class of Fuchsian functions .. I had only to write up the results, which took me a few hours.*
>
> *Henri Poincaré, quoted by Bell (1937)*

The results, but often not the workings themselves, then leak into conscious awareness. They are commonly perceived by the discoverers as creative flashes. On this view such insights come "from nowhere" in the same sense that the blindsight patient's guesses come "from nowhere". Evidently it is necessary for us humans to be able both to sense phenomena and to attend to the world to which they relate. Only then do we have a "somewhere" to put them.

Ramachandran's patient Ellen is unable to attend to anything in her left visual field. She contrasts with Drew in that if her attention is drawn to a left-field object, her intact primary visual cortex allows her to see it. But damage from a stroke to Ellen's right parietal lobe has caused the condition known as "hemineglect." The stored mental map of her world lacks a left half. It is interesting to see what happens when this visualizable world fails to support her logical world. In the latter world, clock faces must have 12 inscribed numbers, as entailed by known facts about the times of day and night.

> *When I asked her to draw a clock, Ellen made a full circle instead of just a half circle. This is a fairly common response because circle drawing is a highly overlearned motor response and the stroke did not compromise it. But when it came time for Ellen to fill in the numbers,*

she stopped, stared hard at the circle and then proceeded to write the numbers 1 to 12, cramped entirely on the right side of the circle!

V.S. Ramachandran and S. Blakeslee, 1998

Two case studies

The Duality Principle will be illustrated in two contexts. In both, agents have performed feats of collaborative discovery unthinkable without inter-agent collaboration. In the first case the collaboration is between members of two castes of ant. In the second it is a three-way collaboration between a chess-master, a programmer and a developing Turing-style "child machine". Each discovery presents features that outrun what an individual human could accomplish.

The first concerns discoveries of fact. The discovery agents were members of various species of Siberian wood ants. Discovery by the ant teams of new food sources are of course limited to "novelties of fact". The only "inventions" are the collaborative mechanisms themselves, designed and built in this case over tens of millions of years of evolution.

In the second, the agents were human-computer complexes and followed A.M. Turing's "child machine" prescription for collaboration between an expert human teacher and a teachable program. The application was to a chess end-game that had previously defeated all attempts to construct a viable theory.

The respective claims made are that:

1. Scout-to-forager signal transmission in structured ant teams use advanced logical and arithmetical operations to communicate food locations. The ants also change their coding system when changed conditions offer incentive to re-optimize. The cognitive component of this collaborative skill is deployed in the scout. Foragers only have to interpret and execute the received instruction-strings. The collaboration achieves what neither could achieve alone.
2. The technique of structured induction used in Turing's child-machine style can instil into a program a completeness and depth of *theoretical understanding* of a complex domain that greatly exceeds that attained by the unaided human intellect. Instilling is not by programming but by *teaching by examples* within a teacher-supplied framework.

Both run strongly counter to the cultural flows of the respective fields of ant biology and of AI. Seismic shifts of perception and practice may be required if their plain message is to be assimilated into either scientific canon.

Case study 1. Symbolic communication in ants

The following account is based on my report to the Royal Society of London on a visit to Siberia during 18^{th} June - 2^{nd} July 1998, jointly supported by the Royal Society and by the Russian Academy of Sciences.

The Siberian work conducted by Dr Zhanna Reznikova forms a structured sequence based upon more than 15 years of systematic field and laboratory studies. Her husband and co-worker, Dr Boris Ryabko, is a mathematician specializing in coding theory. His role in the work has been design, analysis and interpretation. In spite of forfeiting visibility by scanty choice of Western publication outlets, both workers are independently known and internationally respected for sustained studies in their respective fields. Some information theoretic aspects of the work are summarized by the authors in *Complexity,* **2** (2), 37-42, Wiley 1996, ISSN 1076-2787 (see `http://journals.wiley.com/`). References are there given to some of their English-language experimental reports on ant cognition and communication.

The experimental series began in the early 1980's with Reznikova's refutation of a dictum still found in Hoelldobler and Wilson's 1990 treatise *The Ants*, that ants cannot recognise each other and thus cannot form and maintain specialized teams ('cliques' in their terminology). The three species used in these studies base their economy on outdoor foraging work conducted by just such teams. In the Siberian work, indoor colonies are maintained in the laboratory. Every participating ant is individually marked. Each team consists of a scout, who prospects for food hidden in experimentally constructed mazes, together with her foragers who then go and get it. Between these two phases are interposed the following events:

1. return by the scout to the nest to locate and mobilize her own team,
2. a scout-to-foragers communication session using antennae, legs and palps, lasting from tens of seconds to many minutes and
3. an excursion of the team (minus their scout who has meanwhile been experimentally removed) to the food location.

Key move: record the durations

In the months before my Russian trip I studied the authors' ten published English-language papers and questioned them *via* email. Their key move was based on the fact that they knew on each occasion the location of the hidden food, having themselves placed it in a selected end-point of an experimental maze. In the default condition selection of end-points was made at random. In one critical variant, random selection was conditioned on a strong bias towards one, and in some cases two, pre-selected end-points.

The experimenters were thus able to make inferences about message structure from recorded message *durations* paired with the known complexities of their content. The experimenters pre-determine the content (i.e. what information the scout must transmit if her foragers are to go to the right place) through their freedom to construct different topologies of maze and different end-points in the maze to place the food.

In a given experiment records of each of the constituent trials have the form

Trial no., Date, Known location, Message (secs), ID of team.

On her return to the waiting team, there are two types of possible question that the scout's message might be framed to answer, namely:

(1) What must we do to find the food?'

or alternatively

(2)Where is it?'

The possibly subtle-seeming difference between (1) and (2) has ramifications. It is a question of whether message recipients can interpret only command-strings, or relational descriptions as well. This question, although important, was deferred. In the experimental series on which I concentrated, the scout has only to transmit to her team a single integer, corresponding to the ordinal number of the side-turning in a sequence of some 40, at the end of one of which the food lies. In this case a sufficient explanation of observed behaviours can be constructed in terms of (1) alone. Other experiments (mentioned but not reviewed here, *see* 'Co-ordinate grid exams') yielded suggestive evidence that the ants can indeed also form, transmit and use "cognitive maps". In all categories of experiment I satisfied myself that laboratory procedures had excluded obvious pit-falls, such as ant-to-ant communication of directions *via* odour trails.

My two-week visit focussed on the results of experiments designed to explore ant capabilities to encode, to decode, and to change the coding base, of messages each of which essentially encodes a single integer. I do not here include the 'superlong' messages, see later.

Determining message durations entails only the use of a stop-watch, and an objective criterion for recognising the beginnings and ends of transmission. Mathematical cryptography, in particular Shannon's statistical theory of information and Kolmogorov's algorithmic information theory, allows much to be inferred about the means of encoding just from message-lengths alone, - given that one already has a crib, in the sense of having advance knowledge of the contents of messages. By the time of my arrival on 18$^{\text{th}}$ June 1998 I had pared my concerns down to a short list comprising a few pages of handwritten notes, and had conjectured that two weeks on the spot should decisively settle two key points concerning counting.

The first point had essentially already been established by published material, namely 'Does the scout transmit to her team integers in a unary code? ' The answer has repeatedly been shown to be 'Yes'.

The more controversial claim was that, faced with a radical change introduced by the experimenters in the relative frequencies with which particular integers are associated with food, the scouts change their numbering system so as to exploit the new regularities in the interests of message-length economy. The latter is important since transmission rates are of the order of only a couple of bits per minute.

Superlong messages

I had set out on my journey in a mood to reject the investigators' second claim. Their English-language publications contained no hint of events that seemed to me inseparable from the hypothesised change of number code, namely superlong communication sessions intervening between the old-code phase and the new-code phase. Such superlong sessions would surely be needed for each scout to indoctrinate her team in new-code.

In prior email exchanges, Dr. Ryabko had argued against the logical need to expect such tutorial sessions. He did, however, own that, although unreported outside the Russian literature, such unexplained long sessions had indeed been observed. During my stay, I was able with Dr Reznikova's help to follow this up. We combed her original logs and found internally consistent and mutually supportive instances of the postulated superlong sessions. Preliminary analyses of the contexts in which the superlong sessions occur, together with a related phenomenon of 'mass mobilization' also recorded in the logs, convinced me that a *prima facie* case exists for the 'new-code' claim. I was able to bring back photocopies of the most relevant original documents for further analysis.

The new discovery takes the study of animal communication beyond the widely-studied but limited powers of symbolic communication found in the honey bee. For one thing the ants' signalling system embodies a digital as opposed to analogue code. It also calls for a revision of ideas about the generation and transmission of cognitive content, formerly thought to be restricted to higher vertebrates. It seems therefore worth placing the specialized numerical capabilities which I selected for my study visit within the broader experimental framework that the Siberian workers have constructed over the years.

Experimental framework

With the three species that had been shown to have scout-forager teams, all experiments have two parts. The run-up consisted of familiarization trials which sometimes lasted as long as a month. During this period a minority of scouts with poor records of competence are identified and discarded. The run-up is followed by lengthy and rigorous 'final exams'. The investigators claim to have established the following propositions.

1. Binary tree exams

The scout can routinely impart to her team 'left-right-left . . .' directions for routes encompassing up to six forks. Message durations are linearly dependent (except as qualified below) on the Shannon complexity of the message, - e.g. a path involving four left-right choices rates 4 bits.

However, for fixed number of left-right choices, different routes have different human-perceived complexities. Thus LLLLLL (six 'left' turns in a row) should be more briefly expressible than LRLRLR, which in turn should have a shorter encoding than the formless LRRRLL). When message length is fixed, these preference-orderings correlate highly and consistently with the corresponding ant message duration, and indicate a Kolmogorov complexity component.

2. Number exams

In *Formica polyctena* most, but not all, scouts can count and reliably transmit integers up to about 40.

The scout communicates the results to her foragers in a unary code (e.g. a human user might phrase '4' as 'finger finger finger finger').

She can re-optimise her numerical code in response to experimental modification of the statistical properties of the task. The modified codes use simple additions and subtractions, achieving economy in a manner reminiscent of the Roman numeral system when compared with 'finger finger finger ...'

3. Co-ordinate grid exams

A scout who finds food at one of the intersection points of a rectangular grid of stout wires, transmits a message of a form quite different from the 'binary tree' instruction sequences. Here the transmitted information appears to convey the co-ordinates of the distant location. The team goes to the spot, but each by her own, sometimes roundabout, path. It was to these experiments that I referred when speaking earlier of "cogntive maps". Although strongly suggestive, the data are less massive than those of series 1 and 2 above, and lack year-on-year repetitions.

The three species concerned are *Formica sanguinea*, *Formica polyctena* and *Camponotus saxatilis*. Informal trials of *saxatilis* in the numbers game were disappointing. But other observations show that on some tasks they can be as smart as the others. All-round, *F. polyctena* score best.

Points of perspective

Some points of perspective are needed. The observations apply only to a few 'freak' species of ants, i.e. selected from the ten or so that have been described as highly socialized out of a total of some 6000 species. Today's *Homo sapiens* has experience of a densely urban life-style which places a premium on symbolic communication. This experience belongs only to recent, i.e. post-evolutionary, history. Otherwise such lifestyles are confined to the social insects, notably the ants, termites and bees. There is a difference. Experience of complex social existence is known to reach back many tens of millions of years, necessarily if required skills are to be evolved as innate routines. Acquisition *via* genetic evolution is indeed the only path. Their relatively simple nervous systems, coupled with the impossibility in ants of inter-generational cultural transmission, rules out the human trick of inventing and culturally transmitting symbolic communication as required.

To elaborate on the role of nature versus nurture, the numerical and communicative competences found in these species of ants must have arisen in the genetic code over evolutionary time, rather than being re-discovered in each generation by brains innately lacking built-in symbolic skills. For these ants cultural inheritance is not an option. Each new colony is founded afresh by a single fecund queen. Whatever is learned in the previous generation is lost.

Anything resembling cognition in an individual ant has to rest on a rather small neuro-anatomy. The part of an ant's brain (the "mushroom" structure) that is deemed equivalent to the vertebrate cerebral cortex contains a mere quarter of a million neurones. In the human cortex the corresponding number is ten thousand million.

The other salient contrast in this instance between human and ant problem-solving is in the caste-based factorization of a total cognitive task into components. Thus:

- Scouts can translate sensorimotor experience into numerical and logical prescriptions and can transmit the prescriptions to foragers.
- Foragers can interpret the prescriptions and execute them.

Since the period of my visits Dr Reznikova has renewed her attack with a graduate helper on relating the capabilities revealed in the laboratory set-up to behaviours observed in the natural tree-top working environment. Every night the ants return to mound-like nests at the foot of the trees. The following is from an email message recently received:

> *With Tanya Novgoroda we observed antennal contacts by which scouts attracted aphid milkers to the new aphid colony. The scout mobilizes the team from the neighbouring branches of the tree, not from the nest (this would be too far). This suggests a use for symbolic communication: to solve the very important problem of providing the ant colony with honeydew gained from aphids ... the ant colony depends on honeydew to a great extent: they eat more than 1000 kg (in dry sugar weight) per summer. And their success depends on success of scouts which search for new aphid colonies in a very changeable situation (rain, predators and so on).*

Reznikova and Novgoroda also found that having been mobilized by their scout to proceed to a designated leaf to farm a new herd, the foraging team in *Formica polyctena* exhibits further subdivisions of function related to their tasks of aphid-management. Subtasks include mounting guard, herding and milking. It is not yet known whether these roles are exchangeable within the forager group.

Results of extensive tests suggest that foragers cannot communicate symbolically with each other. Technically they are of a different caste from scouts, who are somewhat smaller. In the wild, one scout is responsible for the management of several teams.

Duality Principle again

The hierarchical scout/foraging teams arrangement is not strictly to be compared with the Person-Zombie dichotomy in human cognition. It none the less presents a layered structure of collaborating agents. To that degree it offers suggestive parallels.

Advances are currently being made in continental and British laboratories along several dimensions of ant cognition, including the combination by desert ants of global and local frames in support of navigation by landmarks. Separately from what was described above, Reznikova can reproduce in ants cognitive communications of a kind that raised a stir in the West when first elicited in dolphins (see Andersen (ed.) *The Biology of Marine Mammals*, Academic Press, 1969). With dolphins, communication was peer-to-peer. With Reznikova's ants it was scout-to-forager.

Case study 2. Discovery of a complex theory in chess

An overt version of Turing's proposed child-machine method was applied in the early 1980's to a chess endgame micro-world, namely King and Pawn (on a7) *versus* King and Rook (conventionally written KPa7KR). The program was required to adjudicate, giving its reasoned justifications, claims by the White player (who in this class of positions threatens to promote the pawn to victory) that a given adjourned position is won for White. The resulting hierarchical concept-driven program was subsequently able to demonstrate its chess intelligence by fielding, *and commenting on,* questions put to it at an international gathering of computer chess specialists and chess-masters (*see* Michie, 1986; Michie, 1995a).

In doing so, it immeasurably outran the primitive and sketchy grasp of this end-game which expert human analysis had previously achieved over a century or so of intermittent study. Note that what is here under discussion is *not* competence in correctly attaching the label "win/not-won for White" to arbitrarily presented positions: for KPa7KR, Grandmasters can do this with near-100% accuracy. What is at issue is an agent's ability to produce complete and reasoned justifications for each act of labelling. For KPa7KR, chess-masters are very far indeed from this.

A form of miniature Turing Test was thus found on this occasion to contribute evidence of the method's adequacy for a specified task, namely the inculcation into a computing system of what would be classed as a highly developed form of understanding if manifested by a human.

More than a decade has passed since this first educable child machine was engineered by A. D. Shapiro (1987). He termed the teaching-by-examples method "Structured Induction". The teacher was International Master D. Kopec. Both worked at the Turing Institute, Glasgow. A version of the educable child machine itself was then made commercially available to an industrial client of the Institute. Westinghouse Electric Corporation at Pittsburgh, USA, had been faced with a seemingly intractable problem of optimising, not chess end-game strategy, but automated quality control of a nuclear fuel refining plant.

For the end-game task, generations of end-game scholars had failed to build even the rudiments of a theory. The industrial problem's intellectual complexity proved to be considerably less. A staff member (Leech, 1986) using the identical methodology was able in a few months to teach a closely related related software package to act as a high-performance automatic controller. The consequent marked improvement of yield efficiency brought savings to the company exceeding $10 million per year.

Since then the use of the child-machine software and methodology has permeated a specialized sector of industry. A spot sample of seven case-studies from a single firm is summarized in the Appendix A of Michie (1995b). After Shapiro's monograph went out of print and he had left for a career as a commercial programmer, academic interest lapsed. Recently a partial form of structured induction was revived by an academic group for a pharmacological industry problem. Muggleton, Bryant and Srinivasan (2000) used Inductive Logic Programming (ILP) in something like the same style for discovery of rare structure-activity relations (SARs) in neuropeptides. The estimated reduction of commercial search costs was 1000-fold. The ILP tool used, namely Muggleton's Progol, continues to accumulate chemical knowledge coded in first-order logic for use in computer-aided discovery of further molecular chemical Structure Activity Relations.

Turing's 1950 child machine proposal envisaged precisely such a path, involving what in machine learning is today termed "background knowledge". Turing's prescription (*Mind,* p. 457) was that the educable system should have "a complete system of logical inference built in. The store would be largely occupied with definitions and propositions. The propositions would have various kinds of status, e.g. well-established facts, conjectures, statements given by authority, expressions having the logical form of propositions but not belief value."

In the micro-world of biomolecular theory-discovery, not only is the child machine approach scoring early successes but the spirit of the Turing Test unexpectedly re-appears. The biochemical "teachers" of these prototype systems report that they can only effectively critique the machine-generated molecular conjectures if they can credibly see these conjectures as emanating from a human colleague. To facilitate this the Muggleton group have retrofitted a primitive English-language front end. Turing-style imitation games for assessing the biomolecular intelligence of candidate theory-discovery programs have a foreseeable future place in the pharmacological industry.

Conclusion

The AI research community has in the past viewed intelligence, whether applied to fact discovery or theory discovery, as a unitary system. Just which of two behavioural paradigms is the key - whether Person-like or Zombie-like - has been a topic of dispute. There is an indication in this that a point has been missed. The key, if anywhere, is not in either, but in the interaction. Development of the arts of Zombie-simulation and Person-simulation should be seen as separate preliminaries to integrated collaboration between notionally separable components. The intention of my two case histories has been to illustrate this point. There may be worse places for machine intelligence designers to look than among the varieties of interaction between the brain's layered agent systems and among other layered forms of inter-agent collaboration found in nature.

Finally, it will not have escaped notice that the material reviewed in this paper lends strong further support to Marvin Minsky's (1986) broad philosophy of the human mind as a multi-agent system.

References

Bell, E.T. (1937) *Men of Mathematics.* London: Gollancz.

Galton, F. (1883) *Inquiries into Human Faculty and its Development.* London: Macmillan

Hoelldobler, B. and Wilson, E.O. (1990) *The Ants,* USA: Harvard; Berlin, Heidelberg: Springer-Verlag.

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

Michie, D. (1995b) Problem decomposition and the learning of skills. In *Machine Learning ECML-95,* Lecture notes in Artificial Intelligence, **912,** 17-31. (ed. N. Lavrac and S. Wrobel), Berlin, Heidelberg, New York: Springer Verlag).

Michie, D. (1998) Learning concepts from data. *Expert Systems with Applications,* **15**, 193-204.

Michie, D., Bain, M. and Hayes-Michie, J.E. (1990) Cognitive models from subcognitive skills. In *Knowledge-based Systems in Industrial Control* (eds. M. Grimble, J. McGhee P. Mowforth), Stevenage, UK: Peter Peregrinus, pp. 71-99.

Minsky, M. (1986) *The Society of Mind.* New York: Simon and Schuster.

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

Muggleton, S. and Michie, D. (1996) Machine intelligibility and the duality principle. *BT Technol J*, **14** (4), 15-23.

Ramachandran, V.S. and Blakeslee, S. (1998) *Phantoms in the Brain,* London: Fourth Estate Ltd. (paperback, 1999).

Ryabko, B. and Reznikova, Zh. (1996) Using Shannon entropy and Kolmogorov complexity to study the communicative system and cognitive capacities in ants. *Complexity,* **2** (2), 37-42, John Wiley. Online ISSN: 1099-0526. Print ISSN: 1076-2787.

Shapiro, A.D. (1987) *Structured Induction in Expert Systems.* Addison Wesley.

Turing, A.M. (1950) Computing machinery and intelligence. *Mind,* **59** (236), 433-460. Reprinted 1992 in *Mechanical Intelligence: Collected Works of A.M. Turing* (ed. D.C. Ince), Amsterdam, London, New York, Tokyo: North-Holland.