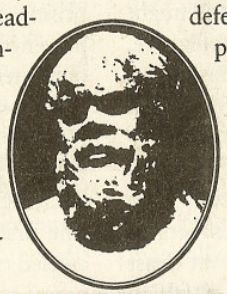
TIM VAN GELDER

INTO THE DEEP BLUE YONDER

THOUSANDS OF TIMES every day, humans pit their wits against machines. Most of the time they lose. Arcade games, bridge programs, chess machines: the phenomenon is so familiar we no longer notice it. We have grown accustomed to being outclassed by electronic gadgets in many activities we find intellectually demanding.

In New York a few months ago, a thirty-four-year-old Azerbaijanie man sat down to a few games of chess against a machine. This event, however, galvanised world attention. Chess enthusiasts followed every move by satellite television or Internet. Newspaper headlines announced the score to millions more. Pundits the world over pontificated on the significance of the occasion.



PHILOSOPHY & IDEAS

Why so much interest in this match? The Azerbaijanie was Gary Kasparov, the reigning world chess champion, widely regarded as the greatest player in the history of the game. Kasparov is so good that very few players in the world today can even give him a serious game. To keep his form up, he likes to take on entire national teams in "clock simultaneous" matches. In these matches, every player – including Kasparov – has at most 2½ hours "thinking time".

The machine was the latest version of IBM's Deep Blue, the most powerful chess computer ever built. The match was billed as the ultimate confrontation of Mankind against The Machine. At stake was more than just Kasparov's personal pride, or IBM's reputation in computer technology. At stake was more than just the title of *best chess player in the known universe*. At stake, apparently, was humanity's self-image as uniquely or supremely intelligent, and hence as entitled to a central or at least special place in the cosmos. At stake also was humanity's place on the ladder of power and authority. Machines with superhuman intelligence might eventually be able to enslave humans in relentless and efficient pursuit of their alien designs. We remain safe only as long as there are at least some white knights like Kasparov, humans still smarter than any machine.

Of course, the score is now a matter of historical record. Deep Blue won the series narrowly, 3½ points to 2½. Fortunately, humanity's spin-doctors already had a face-saving interpretation. Deep Blue, they responded, is a mechanistic *idiot savant*. Kasparov can shrug off his defeat: the match was no more interesting than pitching a pole-vaulter against a helicopter. Humanity can rest content: we are still the smartest beings in the universe; we can still respect our unique intellectual capacities; we are not about to be subjugated by a new generation of ruthless machines.

These, then, are the two main interpretations of the Kasparov-Deep Blue clash. Alarmists see Deep Blue as the vanguard of an approaching army of superhuman intellects. Deflationists see Deep Blue as an overgrown and overhyped cash register. Both interpretations read the confrontation in the context of a world-historical competition between Mankind and The Machine. Alarmists see the match as a pivotal moment, one future historians will designate as the occasion upon which both pride of place and the balance of power were ceded to The Machine. Deflationists insist that The Machine is still stupid and Mankind is still safe.

In fact, both these interpretations are mistaken – or rather, misguided. Any interpretation of what may well be an epochal event is built on a foundation of factual and philosophical assumptions; if these are rotten, the edifice is

Tim van Gelder is currently a Queen Elizabeth II Research Fellow in the Department of Philosophy at the University of Melbourne.

inherently unstable. The situation is even worse when key structural members are fears and fantasies rather than logical implications.

The real significance of the Kasparov defeat is at once more strange and more comforting than either of these simple stories. We are not losing to The Machine, but this is not because The Machine is losing to us. Rather, the very distinction between Mankind and The Machine is under pressure. Long before The Machine could be regarded as having overwhelmed us, it will have *become* us. Ultimately, the loser in this confrontation is not Mankind or The Machine; it is our conception of ourselves as essentially *homo sapiens*.

EARLY IN Stanley Kubrick's famous movie *2001: A Space Odyssey*, the astronaut Dave plays and loses a game of chess against HAL, the spaceship's intelligent onboard computer. This event was taken to demonstrate HAL's intellectual superiority, even more than the fact that it could control the spaceship or converse in normal English. As the plot develops it becomes apparent that HAL is out of control, to the point where it has been killing off human astronauts. Its superior intelligence now makes it a highly dangerous opponent.

HAL is a fictional embodiment of the alarmist interpretation of the Kasparov-Deep Blue confrontation. HAL instantiates what alarmists fear Deep Blue might become: a superhuman, general purpose intelligence, self-interested and pitiless. Standing behind this nightmarish vision is a collection of traditional philosophical ideas. Intelligence is regarded as the operation and outcome of Reason, the ability to make inferences in accordance with the principles of Logic. Reason is a specifically human trait, in the sense that members of the species *Homo sapiens* are uniquely or at least supremely rational. It is Reason, more than anything else, which grants humans a special place in the cosmos; it gives them not only the ability, but also the right and duty to organise the world to their own advantage. Chess is the definitive test of intelligence; the winner is always the one with the greatest ability to apply reason in pursuit of its goals. The best chess player is the most intelligent, and therefore the most rational, powerful and privileged, of all beings.

The letters "HAL" immediately precede the letters "IBM" in the alphabet. Some people believe this is no accident; Kubrick chose those letters in order to highlight the danger IBM and corporations like it pose to humanity. It's a nice story, but it's a myth. The term "HAL" was actually derived from "Heuristically programmed ALgorithmic computer". When Kubrick – who had assistance from IBM in making the movie – found out about the coincidence, he wanted to change the name and was only prevented from doing so by production costs.

Just as IBM would not wish to be linked with the homicidal HAL, so it has tried to dispel the alarmist interpretation of Deep Blue's victory. If Mankind had just been humiliated by the Machine, IBM would have to bear responsibility. Being cast as Dr Frankenstein in the public imagination would hardly benefit their corporate image. For this reason IBM is at the forefront of deflationist counter-reactions to the Deep Blue victory. Despite having invested millions of dollars and dozens of expert-years in the project, they are quick to advertise Deep Blue's limitations. Kasparov, they said, plays with insight, intuition, finesse, imagination. Deep Blue just cranks out the possibilities. According to the IBM counter-hype, the real winners in the Kasparov-Deep Blue confrontation are people like you and me. The RS/6000 SP computer driving Deep Blue will be used in traffic control systems, Internet applications, and a host of other mundane conveniences.

Chess has usually been regarded as the most intellectually challenging game known to man. It would be surprising indeed if a machine could beat the greatest player in history, and yet be fundamentally stupid. *That*, one is tempted to say, *does not compute*. That, however, is the position IBM is taking, and one that was echoed recently by none other than Bill Gates.

Two main lines of thought are used to underpin the interpretation of Deep Blue as harmless *idiot savant*. The first is the idea that Deep Blue's move selection is carried out in an utterly mindless fashion. Whereas Kasparov actually thinks about his options, Deep Blue follows pre-ordained rules specifying vast quantities of simple calculations, none of which require the least bit of understanding. This difference is manifested in the number of possible move sequences the players consider before making their moves. Kasparov, like all human chess players, considers only a few dozen or at most a few hundred sequences. Deep Blue considers literally billions of alternatives in a few seconds.

The fact that Deep Blue can beat Kasparov just shows that brute force can sometimes achieve what would otherwise require real thought.

But if good chess is a matter of selecting the best move, and Deep Blue can examine so many more possibilities, how is it that Kasparov is even in the running? According to this line of thought, intelligence is precisely what

makes the difference. Intelligence is the magic ingredient enabling Kasparov to recognise the overall board situation, to zero in on relevant features, to attend only to the most plausible lines of play, to look far ahead in the game, to be creative and daring in his play, and to learn from his opponent's responses. With none of these abilities, Deep Blue is condemned to witless search of all possibilities, no matter how promising. The fact that Deep Blue can beat Kasparov just shows that brute force can sometimes achieve what would otherwise require real thought.

The second line of support considers Deep Blue's performance in domains other than chess. This argument can be traced all the way back to René Descartes. In his *Discourse on Method*, Descartes considered how one might distinguish a real person from a sophisticated automaton imitating a person. He proposed two tests. The first is that one should attempt to engage the candidate in conversation. A machine, he argued, would never be able to "arrange words differently to reply to the sense of all that is said in its presence, as even the most moronic man can do".

The second test is to explore the range of skills the putative person exhibits. Machines can do certain human-like things exceedingly well – witness the animatronic marvels at a place like Disneyland. However, they can only do those things because they were specifically designed and constructed for the job. Their design precludes them from doing anything else. For example, we now have machines which are better than humans at shearing sheep, but don't expect them to knit a woolly jumper or even make a cup of tea. Humans, by contrast, can do a very wide range of things at least tolerably well. That's because they don't rely on dedicated machinery; rather, they control general purpose hardware (hands, in particular) by means of thought processes.

Descartes believed that the "universal instrument" of Reason is necessary in order to pass both these tests. It is because we can think about the meanings of words that we can hold conversations, and it is because we can think about our actions that we can do so many different kinds of things.

Deep Blue, of course, immediately fails Descartes' tests. It cannot even play checkers, let alone walk the dog or hold a conversation. Deflationists conclude that Deep Blue has exactly zero genuine intelligence, even though it plays the best chess in the world.

THESE DEFLATIONARY arguments certainly undermine the simple alarmist view that Deep Blue is the first of a new generation of superhuman intellects poised to enslave the human race. They do not, however, establish that Deep Blue is a witless moron. More careful consideration of the nature of chess, and the machines which play it, supports the commonsense view that Deep Blue does indeed have at least some measure of intelligence.

Chess is what is known as a formal system. Every board position and every move is well-defined and unambiguous, as are the starting and finishing positions. Further, chess is completely self-contained; nothing outside the board has any relevance to the game. Playing good chess means making a sequence of moves ending in checkmate for the opponent. The hard part, of course, is picking the right move at any given time. The typical number of moves available from any given position is about thirty-five. Whether a move is a good one depends on what the next move of the opponent might be, your response, and so forth. A good player can tell which of these possible sequences of moves and countermoves is advantageous, and hence which of the thirty-five moves to select.

All a chess machine needs to do, then, is to examine all the available move-countermove sequences, and select one ending in checkmate for the opponent. Unfortunately, this simple strategy is completely out of the question (at least, for any technology currently imaginable). The fundamental problem is that of combinatorial explosion. It is illustrated by the following puzzle. Imagine folding a normal sheet of paper in half. The remaining "pile" is twice as thick as the original sheet. Continue until you have folded it 100 times. How thick is the pile now? Most people estimate a few yards. In fact, the pile would stretch eight hundred thousand billion times the distance from the earth to the sun (give or take a few trillion miles).

Combinatorial explosion affects chess just as dramatically. The number of possible move sequences increases exponentially with each "ply" (move), and before long exceeds such familiar measures of enormity as the number of particles in the universe or the number of seconds since the beginning of time. This prevents any conceivable machine from playing good chess simply by mindlessly searching the branching tree of move sequences.

The real secret to good chess is not being able to consider vast quantities of move sequences (though that helps). Rather, the secret is being able to *ignore* the overwhelming majority of sequences, and focus attention on those relatively few which have some real promise. But how do you tell in advance which sequences to ignore? How do you prune from the search tree branches you haven't even looked at?

The answer, basically, is that you use what computer scientists call "heuristics" – rules of thumb providing

reliable, though not infallible guides. For example, a handy rule in finding checkmates is to examine first those moves that permit the opponent the fewest replies. Heuristics are distillations of considerable experience with the domain. At one level, a computer must always be programmed to "blindly" follow algorithms telling it exactly what to do and how to do it. At another level, however, those algorithms can embody heuristics guiding the computer in producing sophisticated – even "thoughtful" – behaviour.

Deep Blue, like all chess computers, operates by means of heuristically-guided search. Its power results from two factors. On one hand, it is an enormously fast search engine. Its 256 specially-designed processors can consider almost a quarter of a billion moves every second; in a game it will examine trillions of possibilities before making a move. On the other hand, and even more importantly, its software embodies a vast amount of real chess knowledge encoded in the form of heuristics. The team of experts who spent years refining Deep Blue's understanding of chess included an international grand master. Almost every match Kasparov has played in the last twenty years has been recorded; Deep Blue is intimately familiar with Kasparov's game.

Therefore, the image of Deep Blue as a prodigiously powerful but essentially stupid "number cruncher" is seriously deficient. Deep Blue embodies a great deal of human-derived chess knowledge, and puts that knowledge to good use in choosing intelligently. Indeed, Deep Blue *has* to be that way; the problem of combinatorial explosion prevents any simple brute-force machine from playing good chess, at least for the foreseeable future.

An interesting consequence is that, as computers have reached the very top levels, their style of play has become more "human". For example, "trappy" moves – ones that gently coax an opponent into an apparent position of strength, but hold a sting many plies down the road – were once a human specialty. These days, with real chess knowledge guiding their search patterns, computers not only avoid traps, they set them. Kasparov himself is no longer able to say, reliably, whether an opponent is human or machine just by looking at the moves. (HAL, by the way, played chess that was quite "human" in style. This was no coincidence; the game in the movie was transcribed from an obscure match played in Hamburg in 1913.)

Deep Blue, then, does have intelligence. It plays a mean game of chess, and does so by thinking about its moves. There are still, to be sure, some significant differences between Kasparov's thought processes and those of Deep Blue. Both, however, are thinking, and the outcome is the same.

Until computer scientists can solve the far more difficult problem of commonsense intelligence, machines will remain our intellectual inferiors and subject to our dominion.

IF THIS IS RIGHT, Descartes' tests cannot be regarded as decisive. There can be genuine intelligence even in the absence of conversation or a wide range of skills. However, Descartes was clearly onto something important. If Deep Blue is so smart, why is it restricted to chess? Why can't it talk about the football?

The deep reason – one of the most important discoveries of cognitive science– is that there are in fact many kinds of intelligence: diverse domains in which intelligence can be achieved, and various ways to achieve it. Some theorists have distinguished as many as seven different categories of intelligence, but the most important distinction for current purposes is that between what we can call *formal* intelligence, on one hand, and *common sense* on the other.

Formal intelligence is that required for domains which, like chess, are formal systems. Such domains might be hugely complex, but they are fundamentally well-defined and self-contained. Common sense is intelligence in domains not satisfying these conditions. Here there is no simple way to specify what the options are, and no way to draw boundaries around what might be relevant. Conversing is the classic example. What do you say when someone says "How are you doing"? Well, that depends – on who said it, in what tone of voice, where they were, what time it was. Try writing a complete set of rules for just the second line of a perfectly ordinary conversation and you'll find out just how much common sense ordinary people actually exhibit.

The difference between formal intelligence and common sense is illustrated by the contrast between formal logic and its informal counterpart. Formal logic is manipulation of symbolic structures in accordance with strict rules. At elementary levels it is a dull, even "mindless" activity (though still a difficult skill for many people to pick up); at advanced levels, it is quite creative. It has been relatively easy to program computers to perform in this domain, though the best logicians are currently still humans.

Informal logic is a matter of determining when somebody is justified in making some assertion. Would further reductions in tariff barriers lead to further unemployment? A great deal of evidence can be brought to bear, but there are no algorithmic

procedures for determining whether the conclusion follows. For centuries, philosophers harboured the misconception that formal and informal logic are, deep down, the same thing – that all informal reasoning is just a complicated version of predicate calculus. More recently it has become apparent that informal logic requires a great deal of "nous", and there is no easy way to translate that into rule-governed symbol manipulation.

Formal intelligence and common sense are both varieties of intelligence; they are both a matter of figuring out what you should do to achieve your goals within a certain domain. However, they are very different, and they do not easily adapt to each other's roles. On one hand, ordinary people have buckets of common sense (well, most of them, most of the time), but they are inept at chess, mathematics and formal logic. On the other hand, formal intelligence doesn't automatically provide common sense. There is of course the stereotype of the absent-minded physics professor. More seriously, Deep Blue can't do the weekly shopping and there is no simple way to adapt its prodigious formal intelligence to that apparently elementary task.

Traditional artificial intelligence – the science and engineering of smart computers – has grappled with both kinds of intelligence. Its success in formal domains has been matched by a notable lack of success at reproducing common sense. The standard approach has been to attempt to translate the informal domain into an approximately commensurate formal system. Unfortunately, this enterprise is extraordinarily difficult, to say the least. There are some research projects around the world grappling with the problem, but don't hold your breath.

From this perspective, Deep Blue's victory does signify something important about artificial intelligence: namely that, as one expert put it, the easy (formal) part is now almost over, and the real work is just beginning. Computers are reaching superiority in a kind of intelligence which is rather difficult for humans to achieve. However, they are barely at first base with regard to the kind of intelligence humans find entirely natural – negotiating their way around the everyday world.

THUS FAR, I have argued that neither the simple alarmist interpretation, nor the simple deflationist reaction, can be sustained. Deep Blue is not a superhuman intellect, but neither is it just a cash-register on steroids. It is an enormously sophisticated machine exhibiting a significant measure of intelligence in one formal domain, and none in all others. Until computer scientists can solve the far more difficult problem of commonsense intelligence, machines will remain our intellectual inferiors and subject to our dominion.

Is this likely to happen, and if so, when? Some philosophers have claimed it will always be impossible for digital computers to exhibit any significant degree of common sense. Hubert Dreyfus of the University of California at Berkeley is the most important of this group. He has provided powerful arguments that common sense depends upon vast quantities of everyday knowledge and knowhow which can never be fully articulated in a form useable by digital computers.

Such predictions, however, are inherently risky, for they depend on our current levels of understanding of the nature of the problem and the limits of technology. Meanwhile, many researchers are tackling various aspects of the problem and making what counts at least as piecemeal progress on the fringes. The most famous and ambitious of these efforts is the "CYC" project pioneered by Doug Lenat. The goal here is to "upload" the entirety of human commonsense knowledge into a vast electronic encyclopedia ready for use by other programs. The CYC people claim to have commercial applications up and running.

My own opinion is that researchers in artificial intelligence will eventually succeed in solving the problem of commonsense intelligence. It will not be anytime soon. Cracking the chess nut took about four decades longer than originally predicted. In the meantime, we have come to understand that chess was the easy problem. Common sense may well take centuries. Alan Turing, the father of artificial intelligence, predicted in 1950 that by the end of the century – that is, by around now – we would have machines able to converse at pretty convincing levels. No such luck. You can, if you like, interact over the Internet with the best "conversation" machines in the world today. The experience is sure to impress upon you the difficulty of programming a computer with common sense. Nevertheless, progress is being made. The goal – genuine intelligence on tap – is so valuable that vast resources and ingenuity will be thrown at it over the next few hundred years. My money, for what it is worth, is on the side of the computer engineers.

In the case of chess, truly excellent levels of play were only achieved once scientists had developed sufficient understanding of how humans manage to play the game so well, and figured out how to transfer some of that understanding into the computer's design. Deep Blue's intelligence was thus largely a matter of human intelligence, abstracted out and reimplemented in digital hardware. The same will be true in the case of common sense. Constructing computers which hold conversations will only be possible once we understand much better what it is that an ordinary person knows, and how that knowledge is organised, accessed and updated.

Once these problems in cognitive science have been solved, the computer scientists will face the challenge of building electronic instantiations of the same principles.

In other words, artificial intelligence succeeds in part through mimicry. It produces silicon simulacra of the basic principles underlying human intelligence. This is because the fundamental requirements of intelligent performance are universal; what varies are their implementations in different kinds of hardware. Evolution developed in humans a neurobiological implementation of the solution to the problem of commonsense intelligence. Artificial intelligence will develop an alternative implementation of what is, at the relevant abstract level, the same solution.

SUPPOSE THIS is correct. Suppose that in fifty years or so computer scientists have succeeded in producing, say, an automatic personal banker. You dial the bank on your videophone and are connected to a virtual "talking head", a kind of supersmooth version of the eighties television character Max Headroom. You interact with this artificial persona much as you would with an ordinary human being. The conversation is quite intimate; your banker has a name, a personality, and knows quite a bit about you from the bank's files and your previous interactions. As long as you don't stray too far from the world of deposits, balances and mortgages, the illusion that you are interacting with a flesh-and-blood human will be overwhelming.

Now for the crucial question – is this personal banker human or machine? More generally, will artificial intelligence be producing artificial humans, or just machine intelligence? At one extreme is the hardline view that nothing can really *be* human unless it is *Homo sapiens*, that is, shares our evolutionary ancestry and our biological incarnation. According to this position, no matter how sophisticated these systems become, they will always be *mere* machines, imitating but never instantiating human nature. At the other extreme is the ultra-liberal view that membership of *Homo sapiens* is at best an accident of history, and has no essential connection to one's ontological, social and ethical status as human. It took many centuries, but in the West at least we finally arrived at the enlightened view that the borders of humankind have nothing to do with those of gender and skin colour. Some people now argue that we should extend these borders further to include dolphins and other putative intelligentsia. The point is that recognition as "one of us", with attendant rights and responsibilities, should depend not on arbitrary details of one's history or embodiment but on one's capacity to participate in human forms of life. Taken to its logical limit, this view would extend the privilege of human status even to suitably programmed computers.

The philosophical choice between hardline biologism and a more catholic liberalism is not an easy one, and I don't intend to adjudicate the matter here. The point of interest is that artificially intelligent machines participating in human forms of life are the kind of case which put pressure on the seemingly simple distinction between Mankind and The Machine. For most of the industrial age, the distinction was obvious enough – people were flesh and blood, born of woman, rational and emotional, social and spiritual. Machines were metal and electricity, born of the workshop, cold and insensitive. The utterly alien character of traditional machines made it easy to see the relationship between Man and Machine as one of opposition and perhaps competition. This attitude of "us against them" is still with us even in the age of information technology. Thus the Kasparov-Deep Blue match is cast as a critical episode in a kind of cosmic struggle to the death between humanity and the machine.

By the time computers have been programmed with common sense, the contrast between Mankind and Machine will have become blurred, if not entirely overthrown. Computers which match our everyday forms of intelligence, and achieve this precisely because they recapitulate the basic principles underlying our own intelligent behaviour, have become very much *like* us. It will not be easy, either psychologically or philosophically, to draw a rigid distinction between people and PCs. Of course, it will always be possible to maintain doggedly that human nature is essentially a matter of lineage or embodiment, and to distribute rights and privileges accordingly. But as philosopher Robert Brandom remarked, "'We' is said in many ways." There is an unavoidable element of arbitrariness in deciding that "we" will stop at the boundary of our species. Many will choose to draw the boundaries somewhat differently, and in the process revise the very concept of humanity.

The march of information technology will not produce machines superior to humans. Rather, it will overhaul our understanding of what we are and what machines are.

I am suggesting that machines will *never* outperform humans in a general intelligence contest. By the time any such confrontation could conceivably come about, the conceptual contrast between human and machine, upon which the apparent

interest of the contest depends, will have been drastically revised. Computers with common sense will not be humans, in the ordinary sense of today. Neither, however, will they be machines, in the ordinary sense of today. They will be a new entrant on the ontological stage, displacing forever the current constellation of concepts in terms of which we contemplate our place in the world. The irresistible onward march of information technology will not produce machines superior to humans. Rather, it will overhaul our understanding of what we are and what machines are. It will replace a binary opposition with a rich spectrum of manifestations of intelligence, and a correspondingly rich range of ways of determining who or what counts as one of "us".

Deep Blue's victory over Kasparov was the first major public triumph of artificial, programmed intelligence over evolved biological intelligence. It was indeed an event of world-historical significance. Not, as the alarmist fears, because it signifies the arrival of alien, artificial intelligences as threats to humanity. Rather, it is significant because it is the first major milestone in a long process of transformation of human self-understanding – and hence human *being*. If we see history in Hegelian terms, as a series of stages in the evolution of the human spirit or self-consciousness, Deep Blue's victory lies at the cusp of a new era. Our own mastery of technology and our level of scientific self-understanding are reaching the stage where we can re-create deep aspects of ourselves in non-biological form. This process will gradually force a dramatic transformation of our *understanding* of what we essentially are, and hence in *what* we essentially are. As Kasparov himself put it: "Maybe the highest triumph for the Creator is to see his creations re-create themselves." ❑

Boom & Gloom

Who's Doing What to Australia?

by

Bob Browning

Economic and cultural elites are changing Australia radically.

Some have never had it so good. But ordinary Australians are being required to face high unemployment, job insecurity, growing inequality, rural community decay, infrastructure and services decline, family impairment and other pressures.

National pride and self esteem are under attack. Accusations of racism and populism abound in the media.

This book examines telling extracts on key issues from policy lobbyists and leading media commentators to help untangle the confusion which special interests and their ideologies impose on public understanding.

ORDER FORM

Boom and Gloom **$22.95. With postage $25.00.**
Orders to:
Canonbury Press, P.O. Box 835, Kew, Vic. 3101
Fax/phone order on (03) 9 853 2023

Name.....

.....

Address.....

.....

Order by fax or phone, and the book with an invoice will be sent asap.
Please make cheques to Canonbury Press.