Intelligent Machines: From Turing to Deep Blue to Watson and Beyond

Bart Selman

Today's Lecture

What is Artificial Intelligence (AI)?

- the components of intelligence
- historical perspective

[in part from CS-4700 intro]

The current frontier

- recent achievements

Challenges ahead:

- what makes AI problems hard?

What is Intelligence?

Intelligence:

- "the capacity to learn and solve problems"
  (Webster dictionary)
- the ability to act rationally

Artificial Intelligence:

- build and understand intelligent entities
- synergy between:
  - philosophy, psychology, and cognitive science
  - computer science and engineering
  - mathematics and physics

philosophy

e.g., foundational issues (can a machine think?), issues of knowledge and believe, mutual knowledge

psychology and cognitive science

e.g., problem solving skills

computer science and engineering

e.g., complexity theory, algorithms, logic and inference, programming languages, and system building.

mathematics and physics

e.g., statistical modeling, continuous mathematics, Markov models, statistical physics, and complex systems.

What's involved in Intelligence?

A) Ability to interact with the real world

- **to perceive, understand, and act**
- **speech recognition and understanding**
- **image understanding (computer vision)**

B) Reasoning and Planning

- **modelling the external world**
- **problem solving, planning, and decision making**
- ***ability to deal with unexpected problems, uncertainties***

C) Learning and Adaptation

We are continuously learning and adapting.

- **We want systems that adapt to us!**

Different Approaches

I Building exact models of human cognition

view from psychology and cognitive science

II Developing methods to match or exceed human performance in certain domains, possibly by very different means.

Examples:

Deep Blue (‘97), Stanley (‘05)

Watson (’11) , and Dr. Fill (‘11).

Our focus is on II (most recent progress).

New goal: Reach top 100 performers in the world.

Issue: The Hardware

The brain

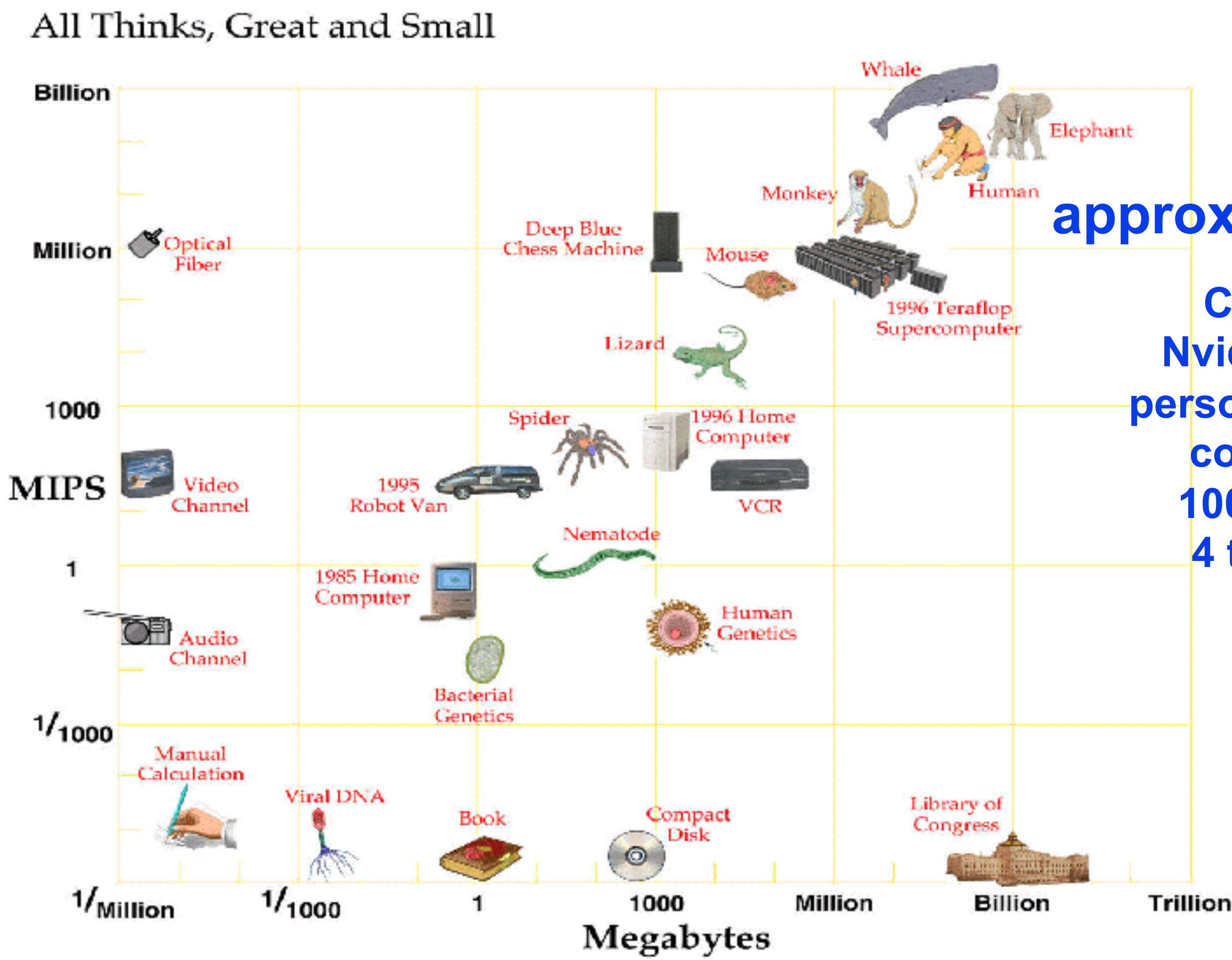
- a neuron, or nerve cell, is the basic information
- processing unit (10^11)
- many more synapses (10^14) connect the neurons
- cycle time: 10^(-3) seconds (1 millisecond)

How complex can we make computers?

- 10^9 or more transistors per CPU
- Ten of thousands of cores, 10^10 bits of RAM
- cycle times: order of 10^(-9) seconds

Numbers are getting close! Hardware will surpass human brain within next 20 yrs.

Computer vs. Brain



approx. 2025

Current:
Nvidia: tesla
personal super-
computer
1000 cores
4 teraflop

Conclusion

- In near future we can have computers with as many processing elements as our brain, but:

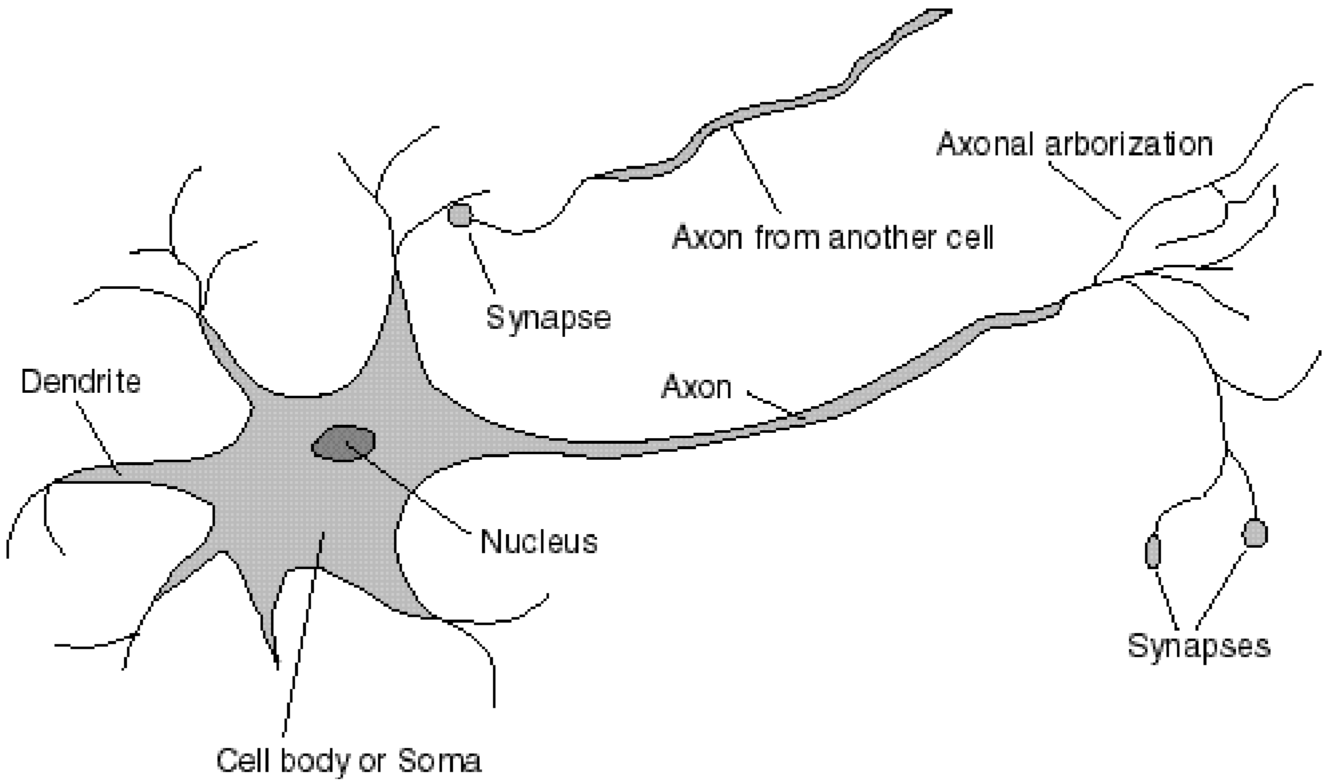
  far fewer interconnections (wires or synapses)

  much faster updates.

Fundamentally different hardware may require fundamentally different algorithms!

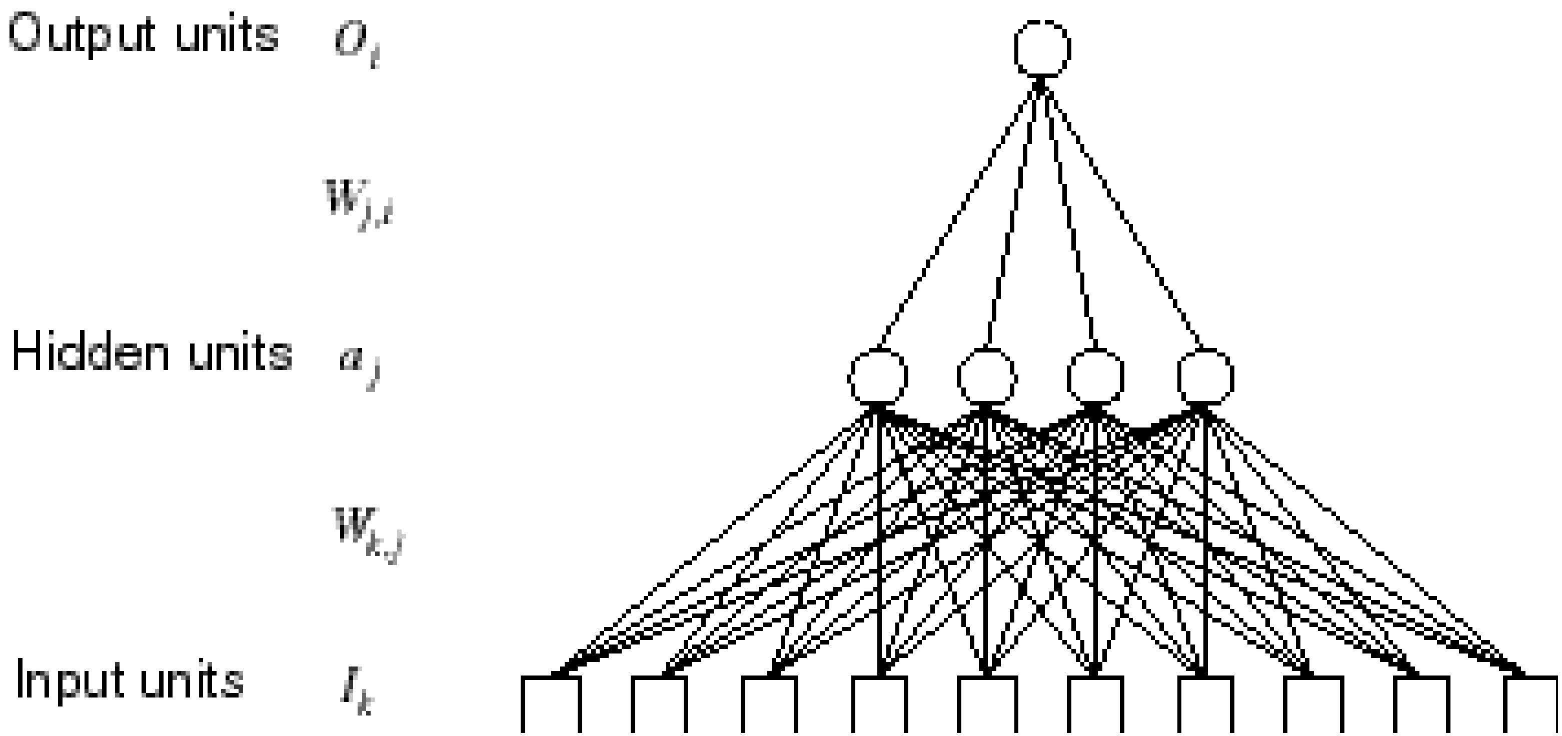
- Very much an open question.
- Neural net research.

A Neuron


Axonal arborization
Axon from another cell
Synapse
Dendrite
Axon
Nucleus
Synapses
Cell body or Soma

An Artificial Neural Network

Output Unit



Input Units

An artificial neural network is an abstraction (well, really, a "drastic simplification") of a real neural network.

Start out with random connection weights on the links between units. Then train from input examples and environment, by changing network weights.

Recent breakthrough: Deep Learning

> **(one of the reading / discussion topics**
>
> **automatic discovery of "deep" features)**

Historical Perspective

Obtaining an understanding of the human mind is one of the final frontiers of modern science.

Founders:

George Boole, Gottlob Frege, and Alfred Tarski

- formalizing the laws of human thought

Alan Turing, John von Neumann, and Claude Shannon

- thinking as computation

John McCarthy, Marvin Minsky, Herbert Simon, and Allen Newell

- the start of the field of AI (1959)

Early success: Deep Blue

May, '97 --- Deep Blue vs. Kasparov. First match won against world-champion. ``intelligent creative'' play.
200 million board positions per second!

Kasparov: "I could feel --- I could smell --- a new kind of intelligence across the table."
... still understood 99.9 of Deep Blue's moves.

Intriguing issue: How does human cognition deal with the search space explosion of chess?

Or how can humans compete with computers at all?? (What does human cognition do?)

Example of reaching top 10 world performers.
Accelerating trend: Stanley (?), Watson, and Dr. Fill.

Deep Blue

An outgrowth of work started by early pioneers, such as, Shannon and McCarthy.

Matches expert level performance, while doing (most likely) something very different from the human expert.

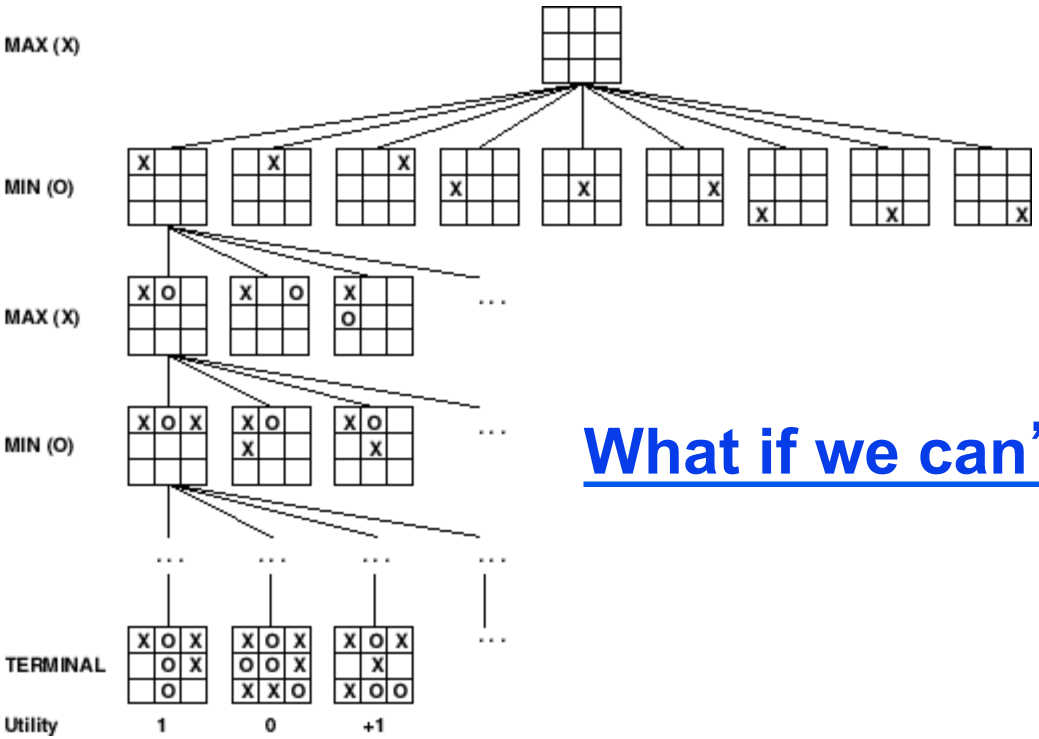
Dominant direction in current research on intelligent machines: we're interested in overall performance.

So far, attempts at incorporating more expert specific chess *knowledge* to prune the search have failed.

What's the problem?

[Room for a project! Can machine learn from watching millions of expert-level chess games?]

Game Tree Search: the Essence of Deep Blue



What if we can't reach bottom?

Aside: Recent new randomized sampling search for Go. (MoGo, 2008)

Combinatorics of Chess

Opening book

Endgame

- database of all 5 piece endgames exists; database of all 6 piece games being built

Middle game

- branching factor of 30 to 40
- $1000^{(d/2)}$ positions
  - 1 move by each player = 1,000
  - 2 moves by each player = 1,000,000
  - 3 moves by each player = 1,000,000,000

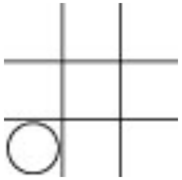
Positions with Smart Pruning

Search Depth		Positions
	Strong player: >= 10K boards	
2	Grandmaster: >= 100K boards	60
4		2,000
6		60,000
8		2,000,000
10	(<1 second DB)	60,000,000
12		2,000,000,000
14	(5 minutes DB)	60,000,000,000
16		2,000,000,000,000

How many lines of play does a grand master consider?

Around 5 to 7 (principal variations)

Why is it so difficult to use real expert chess knowledge?



Example: consider tic-tac-toe. ☺

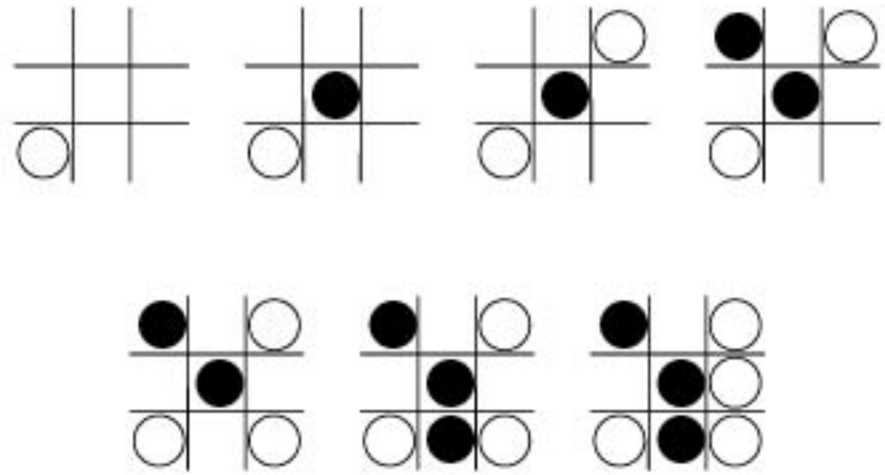
What next for Black?

Suggested strategy:

1) If there is a winning move, make it.
2) If opponent can win at a square by next move, play that move. (“block”)
3) Taking central square is better than others.
4) Taking corners is better than on edges.

Strategy looks pretty good… right?

But:



Black's strategy:
1) If there is a winning move, make it.
2) If opponent can win at a square by next move, play that move. ("block")
3) Taking central square is better than others.
4) Taking corners is better than on edges.

The problem: Interesting play involves the exceptions to the general rules! ☹

On Game 2

(Game 2 - Deep Blue took an early lead. Kasparov resigned, but it turned out he could have forced a draw by perpetual check.)

This was real chess. This was a game any human grandmaster would have been proud of.

Joel Benjamin

grandmaster, member Deep Blue team

Kasparov on Deep Blue

1996: Kasparov Beats Deep Blue

“I could feel --- I could smell --- a new kind of intelligence across the table.”

1997: Deep Blue Beats Kasparov

“Deep Blue hasn't proven anything.”

Formal Complexity of Chess

How hard is chess (formal complexity)?

- Problem: standard complexity theory tells us nothing about finite games!
- Generalizing chess to NxN board: optimal play is PSPACE-hard
- What is the smallest Boolean circuit that plays optimally on a standard 8x8 board?

Fisher: the smallest circuit for a particular 128 bit function would require more gates than there are atoms in the universe.

Game Tree Search

How to search a game tree was independently invented by Shannon (1950) and Turing (1951).

Technique: MiniMax search.

Evaluation function combines material & position.

- Pruning "bad" nodes: doesn't work in practice (why not??)
- Extend "unstable" nodes (e.g. after captures): works well in practice

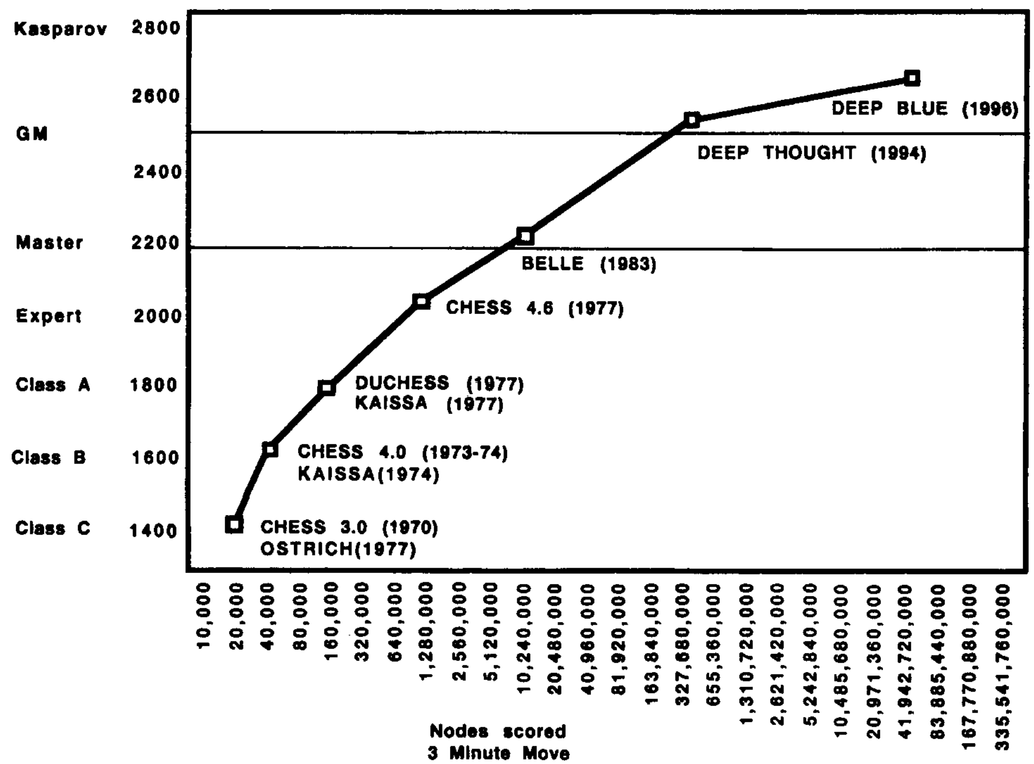


Figure 6.23. Relationship between the level of play by chess programs

A Note on Minimax

Minimax "obviously" correct – but is it?? The deeper we search, the better one plays… Right?

- Nau (1982) discovered pathological game trees

Games where

- evaluation function grows more accurate as it nears the leaves
- but performance is worse the deeper you search!

Clustering

Monte Carlo simulations showed clustering is important

- if winning or losing terminal leaves tend to be clustered, pathologies do not occur
- in chess: a position is "strong" or "weak", rarely completely ambiguous!

But still no completely satisfactory theoretical understanding of why minimax works so well!

History of Search Innovations

Shannon, Turing	Minimax search	1950
Kotok/McCarthy	Alpha-beta pruning	1966
MacHack	Transposition tables	1967
Chess 3.0+	Iterative-deepening	1975
Belle	Special hardware	1978
Cray Blitz	Parallel search	1983
Hitech	Parallel evaluation	1985
Deep Blue	ALL OF THE ABOVE	1997

Evaluation Functions

Primary way knowledge of chess is encoded

- material
- position
  - doubled pawns
  - how constrained position is

Must execute quickly - constant time

- parallel evaluation: allows more complex functions
  - tactics: patterns to recognitize weak positions
  - arbitrarily complicated domain knowledge

Learning better evaluation functions

- Deep Blue learns by tuning weights in its board evaluation function

$$f(p) = w_1 f_1(p) + w_2 f_2(p) + \ldots + w_n f_n(p)$$

- Tune weights to find best least-squares fit with respect to moves actually choosen by grandmasters in 1000+ games.
- The key difference between 1996 and 1997 match!
- Note that Kasparov also trained on

  "computer chess" play.

Open question: Do we even need search?

Transposition Tables

Introduced by Greenblat's Mac Hack (1966)

Basic idea: cacheing

- once a board is evaluated, save in a hash table, avoid re-evaluating.
- called “transposition” tables, because different orderings (transpositions) of the same set of moves can lead to the same board.

Transposition Tables as Learning

Is a form of root learning (memorization).

- positions generalize sequences of moves
- learning on-the-fly
- don't repeat blunders: can't beat the computer twice in a row using same moves!

Deep Blue --- huge transposition tables (100,000,000+), must be carefully managed.

Special-Purpose and Parallel Hardware

Belle (Thompson 1978)

Cray Blitz (1993)

Hitech (1985)

Deep Blue (1987-1996)

- Parallel evaluation: allows more complicated evaluation functions
- Hardest part: coordinating parallel search
- Deep Blue never quite plays the same game, because of "noise" in its hardware!

Deep Blue

Hardware

- 32 general processors
- 220 VSLI chess chips

Overall: 200,000,000 positions per second

- 5 minutes = depth 14

Selective extensions - search deeper at unstable positions

- down to depth 25 !

Evolution of Deep Blue

From 1987 to 1996

- faster chess processors
- port to IBM base machine from Sun
  - Deep Blue's non-Chess hardware is actually quite slow, in integer performance!
- bigger opening and endgame books
- 1996 differed little from 1997 - fixed bugs and tuned evaluation function!
  - After its loss in 1996, people underestimated its strength!

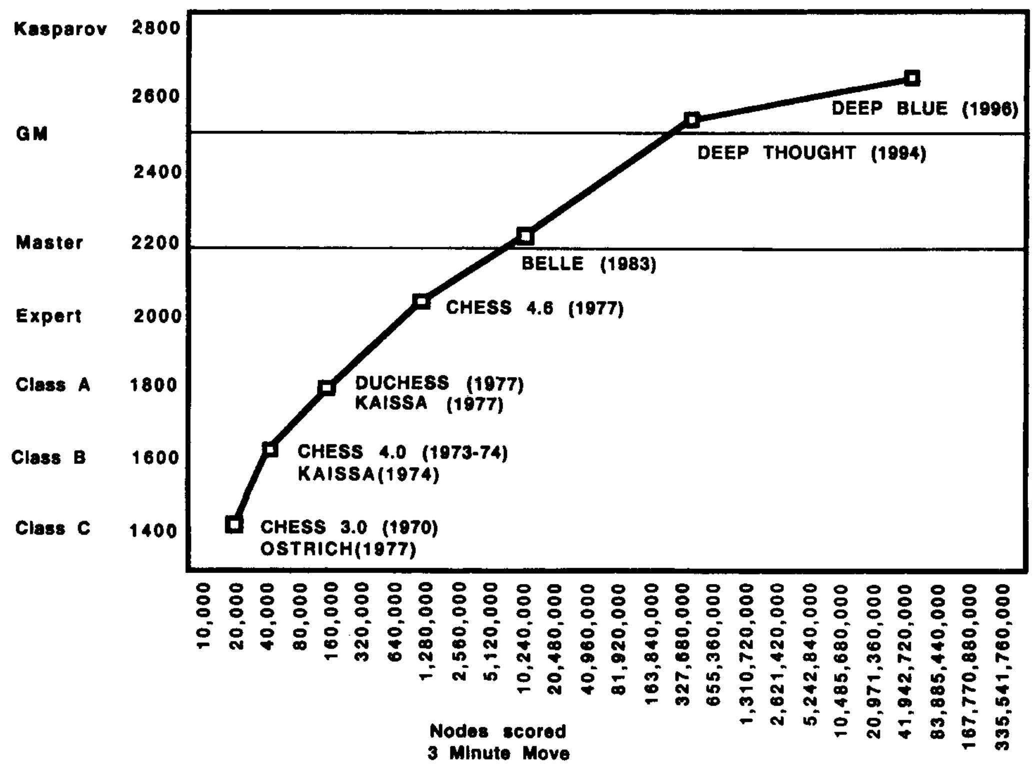


Figure 6.23. Relationship between the level of play by chess programs

Tactics into Strategy

As Deep Blue goes deeper and deeper into a position, it displays elements of strategic understanding. Somewhere out there mere tactics translate into strategy. This is the closet thing I've ever seen to computer intelligence. It's a very weird form of intelligence, but you can feel it. It feels like thinking.

- Frederick Friedel (grandmaster), Newsday, May 9, 1997

One criticism of chess --- it’s complete Information game, in a very well-defined world…

Not hard to extend!

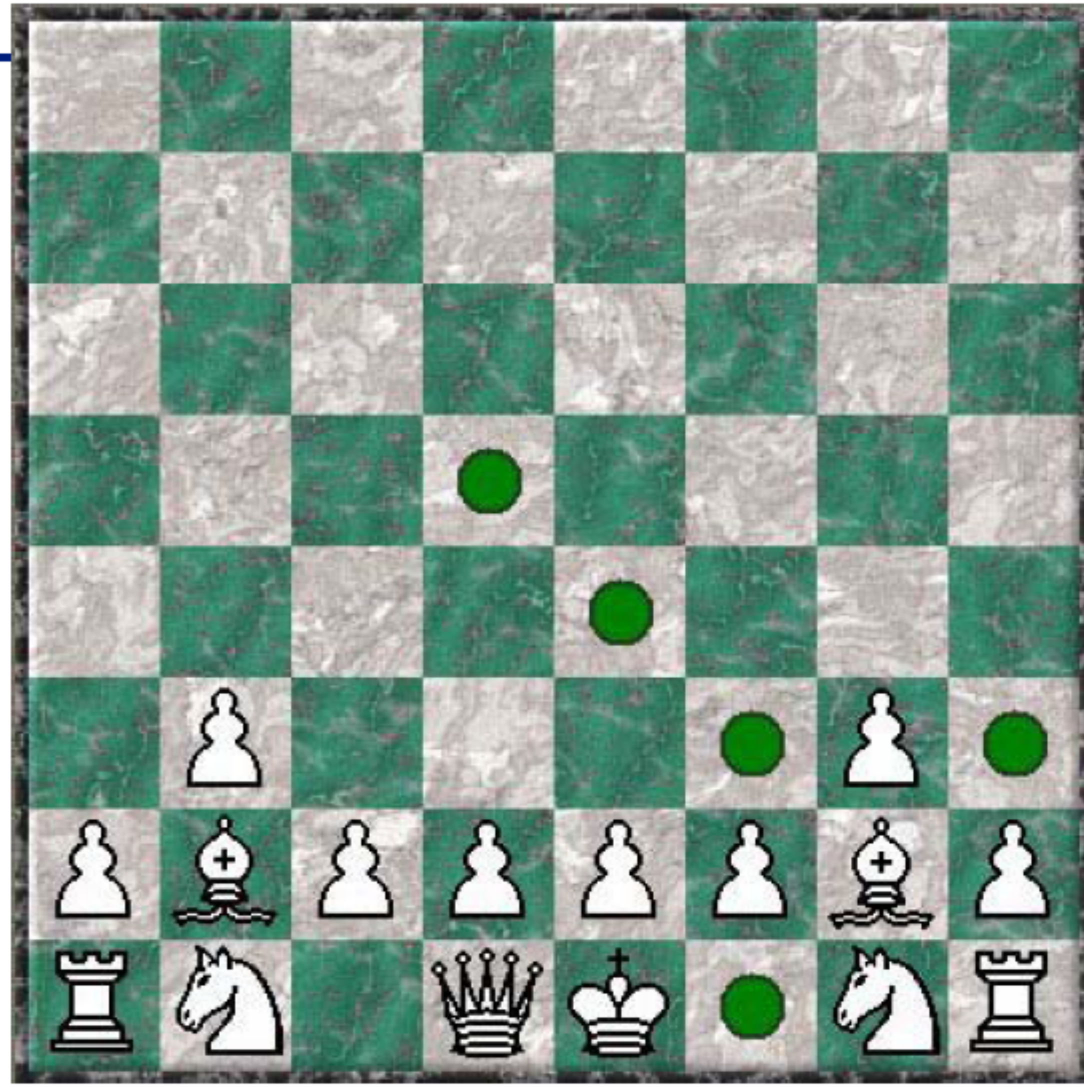
Kriegspiel

Let's make things a bit more challenging…

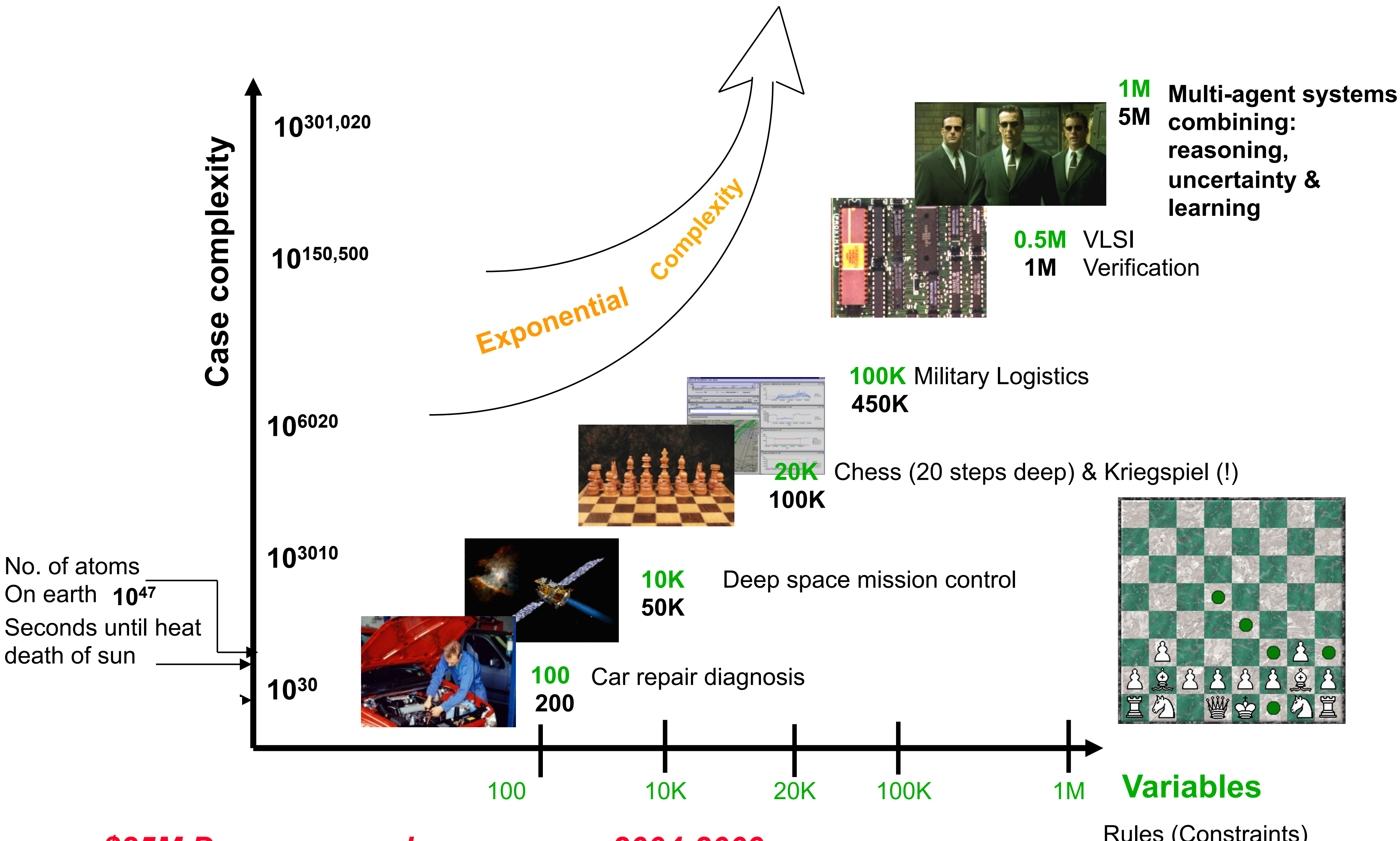
Kriegspiel --- you can't see your opponent!

Incomplete / uncertain information inherent in the game.

Use probabilistic reasoning techniques, e.g., Graphical models, or Markov Logic.



Automated reasoning --- the path



$25M Darpa research program --- 2004-2009

AI Examples, cont.

(Nov., '96) a "creative" proof by computer

- 60 year open problem.
- Robbins' problem in finite algebra.

Qualitative difference from previous results.

- E.g. compare with computer proof of four color theorem.

http://www.mcs.anl.gov/home/mccune/ar/robbins

Does technique generalize?

- Our own expert: Prof. Constable.

ROBBINS CONJECTURE

THE PROOF

7	$\overline{\overline{\overline{p}+q}+\overline{p+q}}=q$	[*Robbins axiom*]
10	$\overline{\overline{\overline{p+q}+\overline{p}+q}+q}=\overline{p+q}$	$[7 \rightarrow 7]$
11	$\overline{\overline{\overline{\overline{p}+q}+p+q}+q}=\overline{\overline{p}+q}$	$[7 \rightarrow 7]$
29	$\overline{\overline{\overline{\overline{p}+q}+p+2q}+\overline{\overline{p}+q}}=q$	$[11 \rightarrow 7]$
54	$\overline{\overline{\overline{\overline{\overline{p}+q}+p+2q}+\overline{\overline{p}+q}+r}+\overline{q+r}}=r$	$[29 \rightarrow 7]$
217	$\overline{\overline{\overline{\overline{\overline{p}+q}+p+2q}+\overline{\overline{p}+q}+\overline{q+r}+r}+r}=\overline{q+r}$	$[54 \rightarrow 7]$
674	$\overline{\overline{\overline{\overline{\overline{\overline{p}+q}+p+2q}+\overline{\overline{p}+q}+\overline{q+r}+r}+r+s}+\overline{\overline{q+r}+s}}=s$	$[217 \rightarrow 7]$
6736	$\overline{\overline{\overline{\overline{3p}+p}+\overline{3p}}+\overline{\overline{3p}+p}+5p}=\overline{\overline{3p}+p}$	$[10 \rightarrow 674]$
8855	$\overline{\overline{\overline{3p}+p}+5p}=\overline{3p}$	$[6736 \rightarrow 7, simp: 54]$
8865	$\overline{\overline{\overline{\overline{3p}+p}+\overline{3p}+2p}+\overline{3p}}=\overline{\overline{3p}+p}+2p$	$[8855 \rightarrow 7]$
8866	$\overline{\overline{\overline{3p}+p}+\overline{3p}}=p$	$[8855 \rightarrow 7, simp: 11]$
8870	$\overline{\overline{\overline{\overline{3p}+p}+\overline{3p}+q}+\overline{p+q}}=q$	$[8866 \rightarrow 7]$
8871	$\overline{\overline{3p}+p}+2p=2p$	$[8865, simp: 8870]$

A Baker's Dozen. *The key steps in proving the Robbins conjecture, as reported by EQP, an automated theorem-proving program developed by William McCune and colleagues at Argonne National Laboratory. (See Box, "Substitute Teacher," page 63 for details.)*

As Eas

The phrase *il est*
repeatedly in the
French mathematician
phrase is common in
spelling out details th
matical techniques.
words. What Laplac
intense mathematical
Oddly enough, La
see how a particular
was easy to see. Th
right combination of
Could a computer
have.
William McCune,
Laboratory in Illinoi
proving programs.
dicts on a range of pr
His most powerful p
can be every bit as
program considers e
the program's creato
Cryptic or not, E
announced a solutio
bolic logic that ha
posed in the 1930's.
has been solved by
described as reasoni
"It's a clear lan
Stanley Burris, a lo
Canada. "Now tha
problem, it opens th
The Robbins co
which are named
(1815–1864), who
into algebraic expr
expressed as $p+$
$p+N(q)$; a logic

NASA: Autonomous Intelligent Systems.

Engine control next generation spacecrafts.
Automatic planning and execution model.
Fast real-time, on-line performance.
Compiled into 2,000 variable logical reasoning problem.

Contrast: current approach customized software with ground control team. (E.g., Mars mission 50 million.)

Machine Learning

In ' 95, TD-Gammon.

World-champion level play by Neural Network that learned from scratch by playing millions and millions of games against itself! (about 4 months of training.)

Has changed human play.

Key open question: Why does this NOT work for, e.g., chess??

Challenges ahead

Note that the examples we discussed so far all involve quite specific tasks.

The systems lack a level of generality and adaptability. They can't easily (if at all) switch context.

Current work on "intelligent agents"
--- integrates various functions (planning, reasoning, learning etc.) in one module
--- goal: to build more flexible / general systems.

A Key Issue

The knowledge-acquisition bottleneck

Lack of general commonsense knowledge.

CYC project (Doug Lenat et al.).

Attempt to encode millions of facts.

New: Wolfram's Alpha knowledge engine

Google's knowledge graph

Reasoning, planning, learning can compensate
to some extent for lack of background knowledge
by deriving information from first principles.

But, presumably, there is a limit to how
far one can take this. (open question)

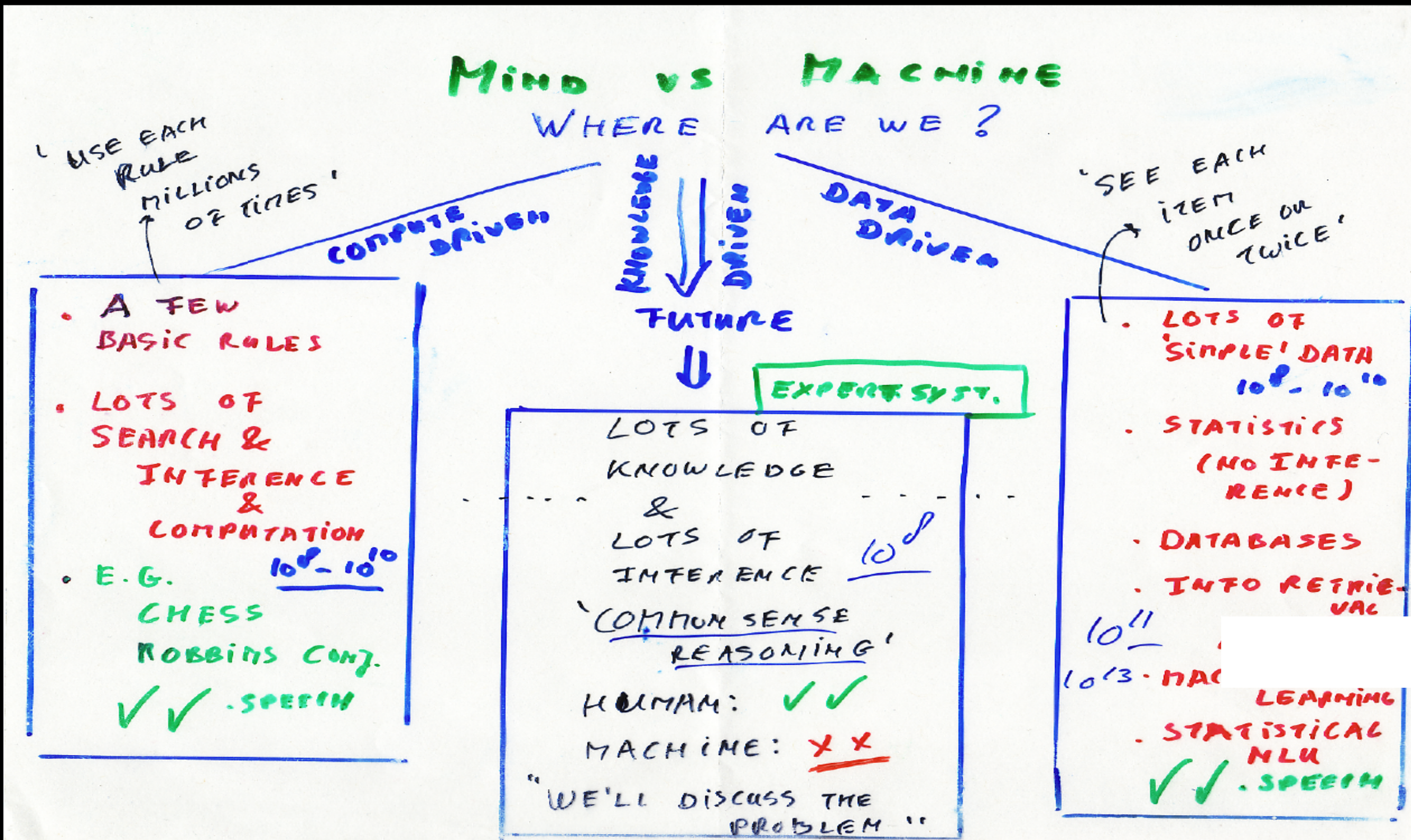
Current key direction in knowledge based systems:

Combine logical (“strict”) inference with probabilistic / Bayesian (“soft”) reasoning.

E.g. Markov Logic (Domingos 2008)

Probabilistic knowledge can be acquired via learning from (noisy/incomplete) data. Great for handling ambiguities!

Logical relations represent hard constraints. E.g., when reasoning about bibliographic reference data, and “author” has to be a “person” and cannot be a location.



☺ ☺

☹ But recent progress!

Knowledge or Data?

Last 5 yrs: New direction.

Combine a few general principles / rules (i.e.
knowledge) with training on a large expert
data set to tune hundreds of model parameters.
Obtain world-expert performance.

Examples:

--- IBM's Watson / Jeopardy

--- Dr. Fill / NYT crosswords

--- Iamus / Classical music composition

Performance: Top 50 or better in the world!

Is this the key to human expert intelligence?

Discussion / readings topic.

END INTRO