

# Superminds



THE STUNNING POWER OF PEOPLE AND COMPUTERS THINKING TOGETHER

THOMAS W. MALONE

The wonderful essence of Tom's book is to imagine how people and computers will interact on a massive scale to create intelligent systems

The real impact of technology will come not only from AI but also from harnessing human minds at hyperscale

## HUMANS + COMPUTERS

Humans plus computers and networks have enormous potential. How can such we creatures as ourselves take advantage of this potential?

How can people and computers be connected so that—collectively—they act more intelligently than any person, group, or computer has ever done before?

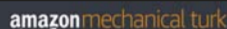
# How Can Computers Help Make Superminds Smarter?

## ROLES OF COMPUTERS?

People have the most control when machines act only as tools, and machines have successively more control as their roles expand to assistants, peers, and, finally, managers.

**Managers?** Even though some people find the idea of a machine as a manager threatening, we already live with them every day. Think about a traffic light, for instance.

For instance, the **CrowdForge** system at Carnegie Mellon University, uses online workers to write documents like encyclopedia articles (..) The average article produced in this way requires 36 separate subtasks, presumably done by 36 different people, and costs only \$3.26 to produce.



**Delegating tasks?** For the Turk workers, this was no different from doing tasks given to them by a human manager, but the automated system allowed the process to be scaled over huge groups of contributors.

## WILL WE EVER HAVE GENERAL AI?

This is like asking, as computer scientist Edsger Dijkstra did, "Can a submarine swim?" Both fish and submarines move through water using their own power, so which word you use to describe that phenomenon is purely a matter of semantics. In English, it sounds strange to use the word swim about a submarine, but in Russian, for example, this sounds perfectly natural

This isn't a debate about the facts; it's just a debate about how we want to define the words we use to describe those facts.

## NEUROMORPHIC COMPUTING

It might be much more feasible in practice to create computers that actually, physically, have billions of processors working in parallel. It might then be much easier to program these more brainlike computers to operate in ways more like human brains. (..) The goal being pursued by a number of research groups today, including at IBM, HRL Laboratories, and elsewhere

## GENERAL AI AS A FORM OF COLLECTIVE INTELLIGENCE?

A "society of mind" emerges from the interactions of many smaller "agents," none of which is very intelligent as an individual but all of which, together, create an overall system that is intelligent.

## HUMAN IN THE LOOP

Long before we have general AI, we can create more and more collectively intelligent systems by building societies of mind that include both human and machine agents.

Cyber-human systems where human and machine agents work together on the same problem

We should move from thinking about putting humans in the loop to **putting computers in the group**, "group versatility" or "group adaptability."

# What Are Superminds?

## COLLECTIVE INTELLIGENCE

Collective intelligence—the result of groups of individuals acting together in ways that seem intelligent.

We can consider a group a supermind if we observe the group trying to do something intelligently.

In order to assess an entity's intelligence, then, an observer always has to make assumptions about the entity's goals.

Individuals in a supermind do not need to cooperate with each other or have the same goals.

## SPECIALIZED INTELLIGENCE

The ability to achieve specific goals effectively in a given environment.

## GENERAL INTELLIGENCE

the ability to achieve a wide range of different goals effectively in different environments.

The definition of general intelligence requires an intelligent actor to be not just good at a specific kind of task but also good at learning how to do a wide range of tasks.

## GENERAL COLLECTIVE INTELLIGENCE

General collective intelligence means "group versatility" or "group adaptability."

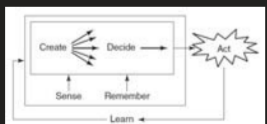
## SUPERINTELLIGENCE

To recognize them I need to identify four things:

- (1) a group of individuals,
- (2) some actions these individuals are taking,
- (3) some interconnections between these actions, and
- (4) some goals with respect to which we can evaluate these actions.

## 3 FACTORS SIGNIFICANT TO THE GROUP'S COLLECTIVE INTELLIGENCE

1. **Social Perceptiveness** - measured by trying to guess the mental state of the person (using a test called Reading the Mind in the Eyes)
2. **The degree to which group members participated about equally in conversation.** When one or two people dominated the conversation, the group was, on average, less intelligent than when participation was more evenly distributed.
3. **Proportion of women in the group.** Groups with a higher proportion of women were more intelligent.



1. In order to act at all, you have to decide what actions to take, even if that decision is made unconsciously.
2. Before you decide on an action, you need to create possibilities for one or more courses of action. But good options for action don't happen in a vacuum. To identify and choose good actions, you almost always need information about the world you're acting in. To get this information, you can:
  - sense the world around you or remember things from the past.
  - 3. Finally, at the heart of intelligence is your ability to learn from experience, to observe patterns in the environment, and to improve your own actions over time.

# How Can Superminds Make Smarter Decisions?

## HIERARCHIES

Since a few people in a hierarchy usually make the most important decisions for the supermind as a whole, their individual human emotions, values, and limitations play a more influential role than in other types of superminds.

**AUTOMATED HIERARCHIES** when machines are doing the routine work that used to be done by people, people will often do new things that were never done before

Example: Site reliability engineering at Google; job is to fix the problems that arise when the machines don't work the way the humans expect them to. For instance, if a program fails because there isn't enough computer memory available for it to run, David needs to figure out why the program ran out of memory and how to keep that from happening again. "If there is something you have to do more than once, then why didn't you write a [program] to do that?" In other words, his job is not just to fix the unusual problems that arise but also to retrain the automated workers so those problems don't happen again.

## FLEXIBLE HIERARCHIES

Adhocracies

## DEMOCRACIES

New information technologies now make it possible to create forms of democracy that combine the best of both direct and representative democracies. These new forms have names like **delegative democracy**, **proxy democracy**, and **smartocracy**.



The architecture of IBM Watson and The Good Judgment Project is similar. Watson includes many different computational agents, each with a different kind of expertise, producing evidence for or against different possible answers. The Good Judgment Project involves a very similar architecture, except that all the agents are people who are connected by computers. **A huge opportunity, I think, is to create "democracies" that include both people and computers.**

Democracies can get smarter by using new technologies to do three things. (1) they can allow voters to express their preferences and values in a much more fine-grained way by delegating to other people or machines the task of voting on their behalf for many more detailed issues than would otherwise be possible. (2) they can facilitate much more accurate ways of determining what is (or will be) true using sophisticated combinations of the judgments of many more people and machines. (3) they may be able to make even more intelligent decisions by separating the first and second functions more clearly.

## MARKETS

the markets in which both people and software bots traded together worked better than those with only people or only bots trading. The combined markets turned out to be significantly more accurate overall, and less susceptible to various kinds of errors. In the combined markets, the people and the computers were all acting as peers, trading with one another, but they each had different kinds of capabilities.

The computers were less likely to be distracted by the specific features of a given situation, more able to systematically apply statistical methods to minimize their errors, and less likely to be overconfident in their own judgments.

The people on the other hand had access to more information than the bots did, and they were better able to respond sensibly to different situations.

## COMMUNITIES

IT can play an important role in both bringing communities together and splitting them apart.

Online discussions today often include lots of repetitions, digressions, and people talking past each other or just ignoring each other. Based on philosophers' notions of how to summarize the essential logical structure of arguments, online argumentation helps reduce the random and disorganized nature of many of today's online discussions

# How Can Superminds Create More Intelligently?

## SPECIALIZED INTELLIGENCE

The ability to achieve specific goals effectively in a given environment.

## MORE PARTICIPANTS

It certainly seems plausible that increasing the number of elementary processing units in a system can help make the system smarter.



The goal of this platform is to crowdsource the process of finding solutions for one of the most important problems facing humanity today: global climate change.

By early 2018, the community had over **100,000** people, including some of the world's leading experts on climate change. (..) together, these people have developed and evaluated over **2,000** proposals.

If these parts aren't connected properly, the system will have no intelligence at all.

## THE BRUTE-FORCE EFFECT

DARPA Red Balloon Challenge A team from MIT by cleverly motivating everyone to not only look for balloons but also recruit others to the cause, this approach rapidly recruited a vast army of searchers. (..) This group's brute-force searching solved the problem far faster than almost anyone expected.



a set of online games that involved folding protein molecules. Thousands of people tried the games. (..) one of its greatest successes, the Foldit community uncovered the structure of an enzyme related to AIDS that had eluded scientists for 15 years. It only took the Foldit community three weeks!

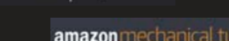
## LIMITS OF SIZE

large groups usually need to expend more effort coordinating their own activities, not just doing the actual work. Even when they make this extra effort, the difficulties of working together may outweigh the benefits of having more people.

## HYPERSPECIALIZATION

Much of the prosperity we enjoy today you might even say much of the collective intelligence of today's organizations - results from specialization of work.

The division of labor that Adam Smith chronicled was for **physical tasks**. I think new information technologies are going to enable another huge wave of division of labor, not for physical work this time but for **information work**.



Amazon originally created the service to help it proofread the product listings on its own website. (..) so some of the workers (called Turkers) become hyperspecialists in particular kinds of microtasks.

there will also need to be workers who specialize in being generalists—whose hyperspecialty is coordinating the work of other hyperspecialists. These workers will perform some of the functions that managers do in today's hierarchies, but their roles may be much more circumscribed.

By reducing the costs of communication and coordination, new technologies are making it possible for groups of people with very diverse kinds of knowledge and problem-solving approaches to work together at a scale and with a degree of collaboration that has never before been possible in all of human history

There are certainly places where hyperspecialization isn't appropriate. In designing a radically new product, for example, the members of a design team may need to develop a shared understanding among themselves about what the product will do before they can think about delegating parts of the design to hyperspecialists.

## COGNITIVE DIVERSITY

Verbalizers are good at reasoning with words; object visualizers are good at dealing with the overall properties of images (like paintings); spatial visualizers are good at analyzing images part by part (as in an architectural blueprint).

The most collectively intelligent groups were those with an intermediate level of cognitive diversity. In other words, groups where the members had very different cognitive styles weren't as smart, perhaps because they couldn't communicate effectively with one another.

# How Else Can Superminds Think More Intelligently?

## CYBER-HUMAN SENSING

People won't begin to be able to analyze this vast amount of data by themselves. But machines won't be able to do it alone. (..) One possible way to do this is with a system that mirrors the multilayered networks of neurons our brains use to interpret the sensory information they receive from the world.

we will be able to combine human and

limitation of the neural-net approach is that even if machines could learn some of these patterns, people wouldn't usually be able to understand how the machines reached their conclusions.

"I predicted this attack because the following 23 million connections in my database had the following 23 million values between 0 and 1."

Today we wouldn't think of trying to run a serious company without a consolidated accounting system that keeps track of the combined results of all the company's financial transactions and packages this information for the people in the company who need to know about it. In the future, it may become just as unthinkable to run a serious organization that doesn't do something similar with many other—much more subjective—kinds of information. In fact, our great-grandchildren may find it hard to understand how the organizations we belong to in the early 21st century could have made so many of their decisions with their eyes—figuratively—closed.

## SMARTER REMEMBERING

Whenever you, as an individual, remember something, you must somehow encode it, store it, and later retrieve it. When superminds remember, they have to perform these three functions, too.



A key aspect of collective memory, as opposed to individual memory, is that collective memory usually requires communication between individuals.

WRITING Starting around 5,000 years ago, this first major information technology profoundly transformed the ways human groups remember.

KNOWLEDGE MANAGEMENT "one of the important lessons researchers learned about knowledge management in the 1990s was that often the most useful thing an online knowledge management system can do is help you find the people who have the information you need rather than the documents with that information."

SMARTER REMEMBERING It's clear that information technology can increase the size and reliability of a supermind's long-term memory. But what about working memory?

When you write items on a whiteboard, they remain visible and accessible to people in the room. That way, everyone can continue to compare and combine the different items on the board without having to remember them all.

Information technology can perform a similar function for much larger groups whose members aren't all in the same room.

Online Argumentation Systems Climate CoLab Human Diagnosis Project (Human DX)



According to the World Health Organization, there are only about 70,000 known human diseases, and a mere 10 of these accounted for 30 percent of all US hospital stays in 2010. (..)

Diagnosing a perplexing case—perhaps of an extremely rare disease—would often just be a matter of Human DX remembering other patients with similar symptoms and what their eventual diagnoses were.

## SPECIALIZED INTELLIGENCE

The ability to achieve specific goals effectively in a given environment.

## CONDITIONS FOR SUCCESS

conditions that would have to be true for each choice to be successful (based on P&G strategic-planning process). A key part of the process was doing research to gauge whether these conditions were true

Michael Porter, who pioneered the modern academic field of corporate strategy, articulated three generic strategies that companies in almost any industry can use: **cost leadership** (being the low-cost producer), **differentiation** (being unique on dimensions, like quality, that customers value), and **focus** (tailoring products to a narrow segment of customers)

Software can help by:
- helping remember good solutions
- suggest combinations of possibilities
- involve more people and tools in generating possibilities (specialists and outsiders)
- use an Evaluation Funnel
- use Prediction Markets, Online Argumentation, and Voting.

## CYBER-HUMAN STRATEGY

The process could start with human experts doing evaluations manually and then gradually automating more and more of the work over time as the machines get better at predicting what human experts would do. (..) I expect that today's strategy consulting firms (or their future competitors) will provide much of this functionality as a service. It is a kind of ecosystem supermind that combines markets (the contests), communities (for proposing ideas), democracies, and hierarchies (for evaluating ideas). And it lets a large group of people and machines apply many more kinds of knowledge and explore many more possibilities than traditional hierarchical problem-solving processes do.

## SMARTER LEARNING

Hermann Ebbinghaus used the concept of a learning curve to analyze how fast you improve by repeatedly practicing something

When you learn by EXPLOITATION, you keep doing the same kind of activity over and over, improving over time

When you learn by EXPLORATION exploration on the other hand, you keep trying new things to see what works and what doesn't, then you do more of the things that work

## CYBER-HUMAN LEARNING

One way to do this is by creating cyber-human learning loops, in which people and computers work together and get better and better over time, often by letting the computers do more and more of the work



Cyber-human learning loops will let machines watch humans prepare more complex returns and gradually learn what actions the humans take in different situations. At first the machines may just suggest actions to humans. (..) Eventually the machines can just automatically take the actions that humans always approve.

I think it will be a very long time before computers will have enough general intelligence to do everything that humans do, but in the meantime, superminds made up of people and computers will be able to continuously improve their performance by learning from their own experiences extremely effectively.

## LEARNING FROM EXPERTS

Computers learning from experts - examples: Adam the robot scientist developed by Ross King and his colleagues at Aberystwyth University and the University of Cambridge. Adam has the intelligence and the physical capabilities to carry out the whole scientific process: originating hypotheses, devising experiments, running those experiments, interpreting the results, and then formulating new hypotheses. Adam's speciality is understanding the genome of baker's yeast, a popular laboratory species. (..) Adam is the first machine to "independently discover new scientific knowledge."

# How Can Superminds Help Solve Our Problems?

## CLIMATE CHANGE

most of the costs of high emissions today will be paid by people who live on our planet in future decades, so most of those costs aren't included in the prices for the things we buy today. What's worse, the benefits of any investments I voluntarily make in reducing emissions (like changing my lightbulbs) are averaged over everyone else on earth, and I won't personally get any noticeable benefit.

A classic example of what economists call an externality. Since these costs are not included in today's markets, markets essentially ignore them when they make their decisions.

## CONSENSUS DECISION MAKING

The basic idea would be to use something like the strategic-planning system focused on climate-change strategies for the world.

In signing the Paris Agreement, almost every country in the world specified voluntary goals for emission reductions and other climate-related actions. If a country doesn't meet its goals, it will be subject to disapproval. (..) to really evaluate whether all these proposed actions are realistic, we also need to gauge whether they are realistic in technical, economic, political, and other ways.

We could use **contest webs**, augmented by a network of people and computers with many different kinds of expertise, to evaluate the overall

## RISKS OF AI

In the 1960s, as computers eliminated large numbers of clerical jobs in the back offices of banks and insurance companies, President Lyndon Johnson created a National Commission on Technology, Automation, and Economic Progress to study the problem.

In the 2010s, Erik Brynjolfsson and Andrew McAfee have warned about the risks of artificial intelligence putting many humans out of work, not just in blue-collar and clerical jobs but in white-collar jobs, too.

## JOBS OF THE FUTURE?

We can make some good guesses based on three observations about how market superminds are likely to decide to use human labor to give us what we want.

• People will do more of the things that machines make cheaper to do.

• People will do things just because we want people to do them.

...

If some individual or small group were able to create an AI in secret much sooner than the community expected, then Bostrom's worst fears might be realized. Is this a theoretical possibility? Yes. But it seems to me that this very theoretical risk is substantially smaller than several other existential threats that humanity faces today

• People will do what machines can't. For the foreseeable future, this includes three important capabilities:

- general intelligence (not just the various kinds of specialized intelligence that machines will more easily master)
- interpersonal skills (beyond the simple ones machines will have), and
- certain physical skills (like operating in unpredictable environments).

McKinsey consulting firm estimates that even though about 50 percent of the activities people are paid to do could be automated with current technology, only about 5 percent of today's occupations could be fully replaced

Interpersonal skills may be even more important than we think. The people who succeed in life often seem to be the ones with the most social intelligence, not necessarily the most cognitive intelligence.

In 1970, automated teller machines (ATMs) were introduced, and by 2010, there were about 400,000 of them in the United States. You might think that the number of bank-teller jobs in that period would have gone down, but in fact it increased slightly, from 500,000 to 550,000 over the years between 1980 and 2010. Of course, there were other things (like bank deregulation and changes in transaction volume) going on during that period, but one of the most interesting explanations for this result is that bank tellers moved from being the equivalent of checkout clerks who just gave people their money to being salespeople who built relationships with customers.

JAMES BESSIN "Foil and Technology" Finance and Development 52, no.1 "Why Are There Still So Many Jobs?"

## SAI'S TAKING OVER

Superintelligent AIs (which Nick Bostrom call SAIs) could plausibly control vast resources. With security lapses and help from friendly humans, it's certainly possible that SAIs could gain effective control.



R.U.R.: Rossum's Universal Robots—the 1920 Czech play that introduced the word robot to the English language

Preventing SAI from taking over the world:

(1) since the individuals in a pure ecosystem have no way of cooperating to control an SAI that becomes more powerful than they are, ecosystems are not a good approach for reducing these risks.

(2) communities could try to reduce the threat of a rogue SAI, their enforcement mechanisms are almost certainly too weak

(3) democracies, like communities, could agree to try to stop SAIs, but their enforcement capabilities are even more limited than those of communities. It certainly wouldn't make sense for citizens to vote individually on each possible decision.

That leaves us only one possibility: hierarchies. Hierarchies can decide to prevent dangerous AI work, and they

## What can we do in the meantime?

Legal responsibility for actions by automated systems. One approach to thinking about these issues would be to use the way current US law deals with liability for the actions of animals. If you own a domesticated animal, like a dog, you are only liable for damages caused by the dog when they result from your own negligence. But if you own a dangerous wild animal, like a tiger, you are liable for whatever the tiger does, even if you haven't been negligent at all.

Human-computer synthesis. Before deploying any automated systems, we should think hard about what could go wrong and how human intelligence could help deal with these situations. We should carefully consider whether and how humans can override the overrides.

...

If some individual or small group were able to create an AI in secret much sooner than the community expected, then Bostrom's worst fears might be realized. Is this a theoretical possibility? Yes. But it seems to me that this very theoretical risk is substantially smaller than several other existential threats that humanity faces today

## People will do things just because we want people to do them

we will pay humans to do some of the things that machines could do more cheaply just because we will want people to do them instead of machines.

we will pay humans to do some of the things that machines could do more cheaply just because we will want people to do them instead of machines. So if all the physical things we need become much cheaper, we'll find other kinds of status symbols to compete about.

## People will do more of the things that machines make cheaper to do

Since many of these jobs don't require full-time work for a single customer, many people will do them as independent contractors. In a 1997 Harvard Business Review article, my colleague Rob Laubacher and I coined the term e-lancers—short for "electronically connected freelancers"

Today many people use the term gig economy to describe essentially the same phenomenon.

# Where Are We Headed?

## CONSCIOUSNESS

Here's a representative list of five of the most common ways of defining consciousness:

- (1) Awareness: an entity is conscious if it reacts to stimuli in the world. Taken broadly, this definition would allow you to say that all living things are conscious because they react to environmental changes.
- (2) Self-awareness: an entity is conscious if it reacts to—and can tell others about—changes in itself.
- (3) Goal-directed behavior: an entity is conscious if it takes intentional action to achieve goals.
- (4) Integrated information: an entity is conscious if it integrates many kinds of information.
- (5) Experience: an entity is conscious if there is something "it is like" to be that entity.

Explaining how and why any system could have a subjective experience is what the philosopher David Chalmers calls the **hard problem of consciousness**.

## CAN GROUPS BE CONSCIOUS?

the answer to the question of whether groups can be conscious is almost certainly yes.

## Is Apple Conscious?

You may be thinking that all these different kinds of responses Apple makes to the world are really the work of individual people. First, even though individual people are certainly involved in these actions, none could have happened without the rest of the group. Second, and even more important, this line of reasoning would lead you to conclude that your own brain is not really conscious—only the individual neurons in it.

The supermind called Apple is certainly aware, self-aware, goal-directed and integrated incompleteness—not just trivial ways. And depending on how empathetic we're willing to be, it seems likely that Apple may even have experiences analogous to the experiences we humans have. It seems that we are justified in saying that Apple is conscious in a way that is closer to being literally true than just metaphorical.

in a certain sense, the global mind existed at least 3.5 billion years ago, when colonies of bacteria developed a kind of intelligent division of labor among their different cells.