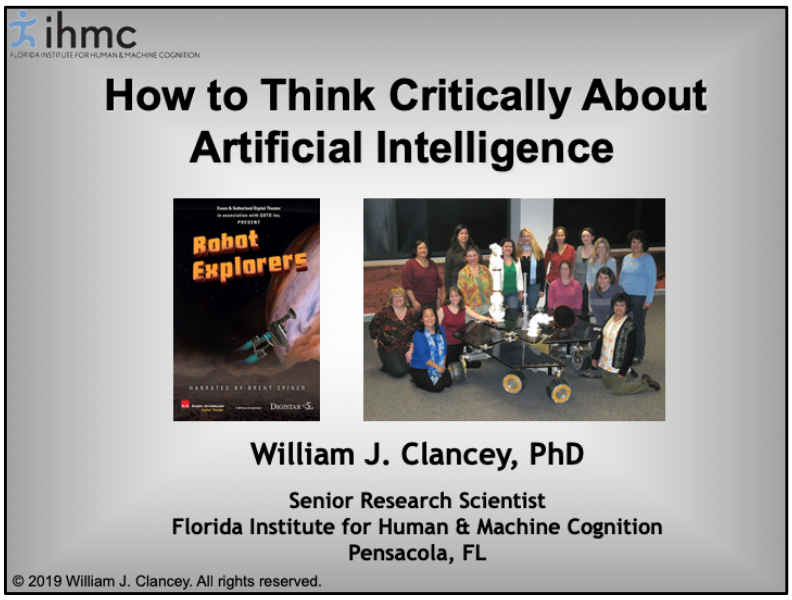
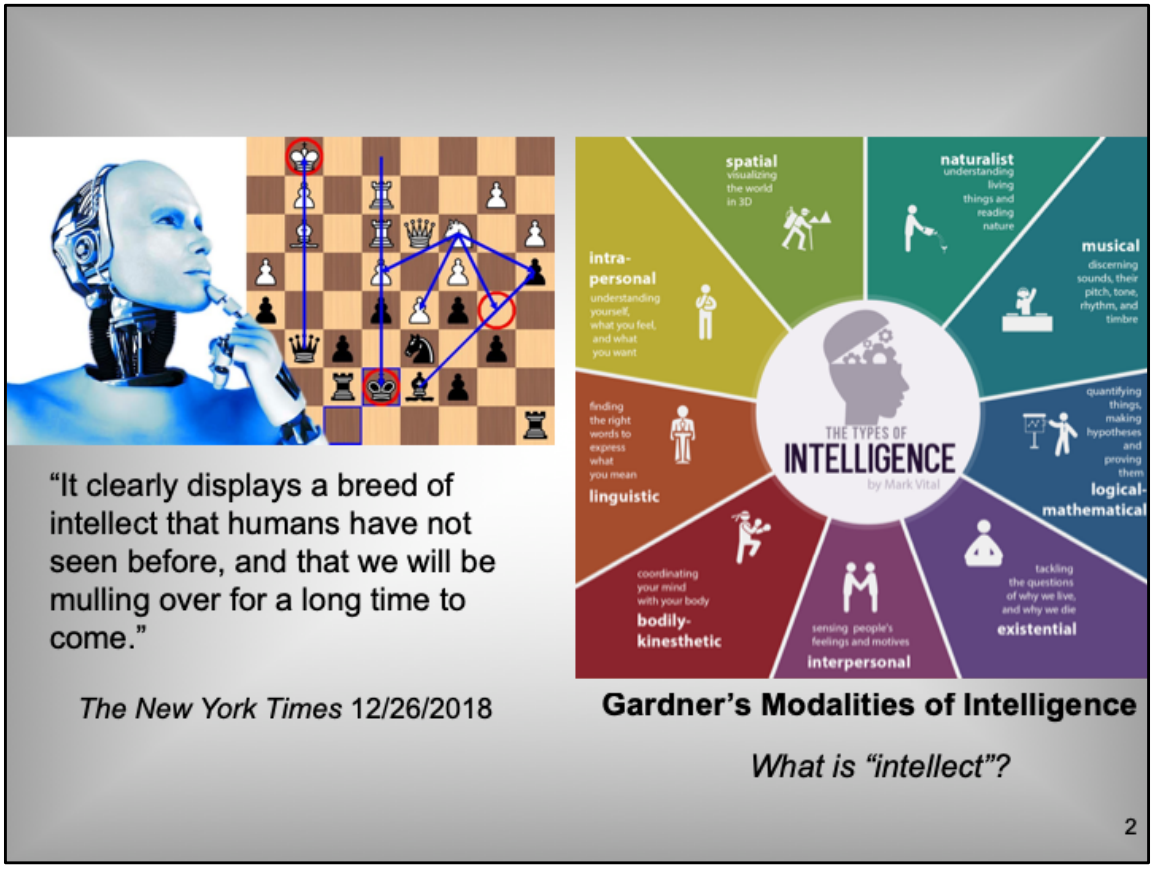


To use AI technology appropriately, improve it, and predict its implications for society requires a scientific understanding of the relation between people and computer systems. When we anthropomorphize automated systems—describing them as exploring, judging, collaborating, etc.—we are claiming from the start what we have set out to do, as if we already have "expert systems" and "smart phones." To describe what computer systems do and how they work, we must be mindful of our words; and if we are to appraise what we have accomplished and how to design systems that fit and complement how people think and work, we must develop better neuropsychological theories of cognition. Accordingly, in this presentation I share my reactions to what people say and write about AI to convey how to think critically about AI research and what language would be more appropriate.

I am focusing on scientific writing and judgments including in the popular press, not poetry, science fiction, or entertaining movies. They all have their place. The problem is that, in journalism and even in AI publications, terminology that only properly applies to people is used to describe machines, possibly to attract attention as a form of marketing or to simplify the topic to make it easier to understand. But descriptions of AI have become clichés; people use these phrases rotely, without considering their meaning and hence what claims are being made.

Consider the two images on the cover slide: On the left is a cover of a documentary about planetary spacecraft. On the right are the women scientists and engineers who are part of the Mars Exploration Rover project. They explore Mars remotely using robotically controlled instruments. If the spacecraft and planetary landers and rovers are also called "explorers," what shall we call the people who use this machinery to explore the solar system? What do explorers do; what do the robots do? If we have robot explorers then why do we need a human spaceflight program? What is the role of the MER science team? Mindlessly anthropomorphizing makes a jumble of the entire enterprise.



The use of clichés and inappropriate metaphors obscures what AI has accomplished. Consider for example how the *NY Times* reported in December 2018 the latest advance in game-playing programs, AlphaZero.: "One Giant Step for a Chess-Playing Machine: The stunning success of AlphaZero, a deep-learning algorithm, heralds a new age of insight — one that, for humans, may not last long." https://nyti.ms/2Rjetd3

What is a "breed of intellect"? Is it like a breed of horses or tulips? How do the computer programs in AlphaZero relate to the intellectual capabilities of people? What is meant by "intellect" in this sentence? This is sensationalism, and of little value for understanding how AlphaZero works or how the technology might be applied (or misapplied).

A better statement would be: "AlphaZero displays champion-level game-playing, using a method for improving play through practice that combines techniques of lookahead search and a memory that records winning lines of play, programming methods that have been developed over the past 50 years of AI research but are combined here in a novel way." Rather than explaining any part of that, the journalist just calls it an "intellect."

Computation is not intellect. What we call intelligence in people has a wide variety of manifestations involving conceptual coordination of verbal, visual, auditory perception with emotion and physical coordination. Gardner described these as different modalities of intelligence (see slide).

(continued on next page)

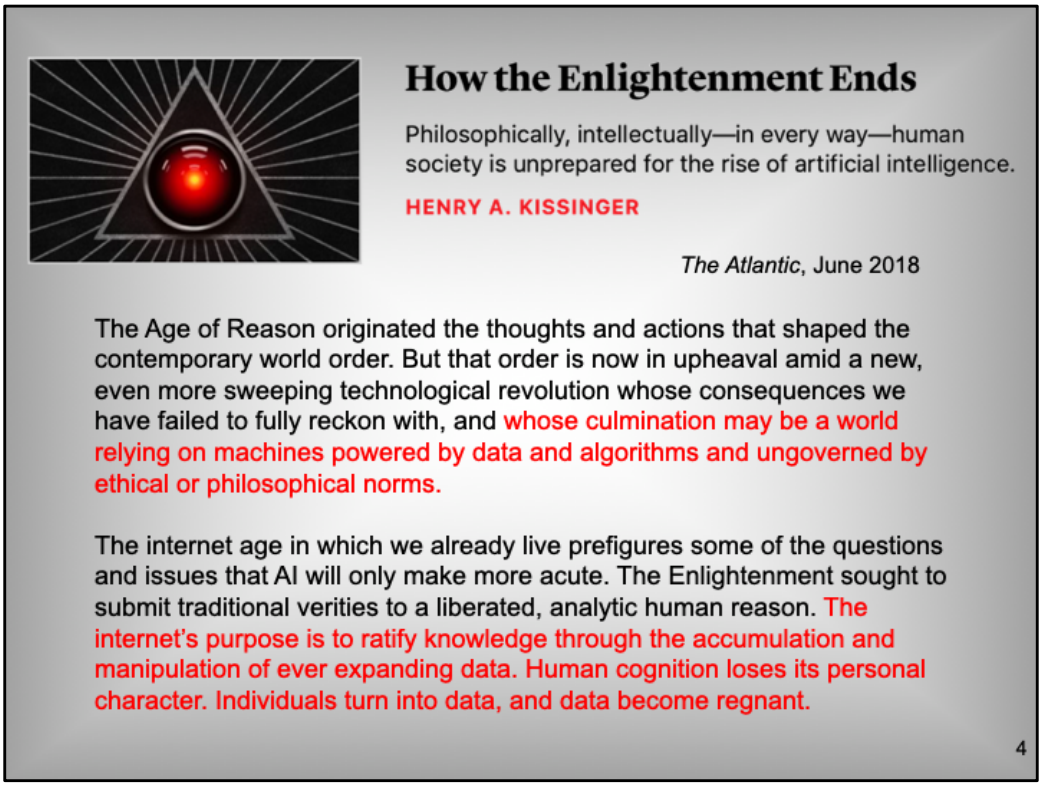
(continued from prior slide)

Playing games with a fixed set of rules in a "closed world" (all information is known and shared) is very different from building a shelter, raising children, or finding one's way in the woods—activities for which a robot that learned using only AlphaZero's methods would not be remotely capable. Its learning program doesn't even apply to competitive bridge, in which the the other players' cards are unknown and bidding follows cultural conventions that cannot be discovered by playing against yourself.

Rather than a "new breed of intellect," the NY Times author could have been "mulling over" how such technology might lead to powerful scientific tools, such as DeepMind's concurrent development of AlphaFold for predicting the 3D structure of proteins from genetic composition. See: https://deepmind.com/blog/alphafold/

It's not difficult to find better articles about AlphaZero, such as Kasparov's remarks prompting us to consider ways people and these tools might work together (Chess, a Drosophila of reasoning. *Science* 07 Dec 2018: Vol. 362, Issue 6419, pp. 1087 DOI: 10.1126/science.aaw2221). It might be advisable to seek out reputable authors, but even professors will have wildly different opinions. The quoted *New York Times* on AlphaZero (labeled as an "essay") was written by Steven Strogatz, an accomplished mathematician at Cornell, who has won awards for writing about science. A very different perspective is offered in an opinion piece by Melanie Mitchell, "While some people are worried about 'superintelligent' A.I., the most dangerous aspect of A.I. systems is that we will trust them too much and give them too much autonomy while not being fully aware of their limitations." See: "Artificial Intelligence Hits the Barrier of Meaning" -- https://www.nytimes.com/2018/11/05/opinion/artificial-intelligence-machine-learning.html

These articles on machine learning show us that the *New York Times*, which many consider to be "a newspaper of record," is not presenting a single, coherent analysis. Rather the reader must sort through and relate multiple "essay" and "opinion" pieces that present very different perspectives.



Henry A. Kissinger served as national-security adviser and secretary of state to Presidents Richard Nixon and Gerald Ford. He is 95 years old (January 2019).

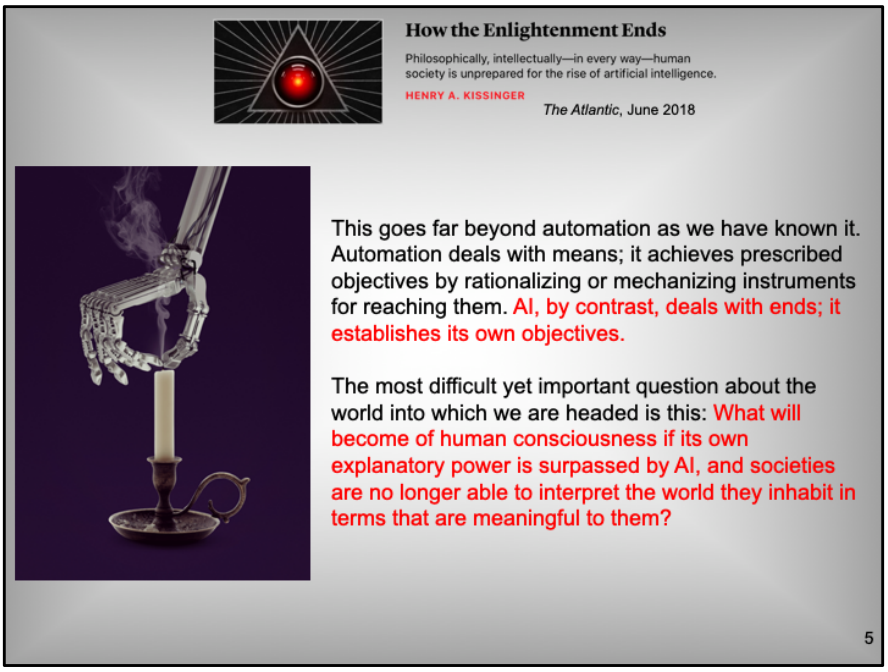
Kissinger's article provides a stark view of a future in which "Ais" dominate our affairs. His writing follows a pattern—a factual or at least reasonably debatable statement about the present is made, following by false claims and hyperbole.

Often it appears Kissinger has a germ of understanding but lacks personal experience to express himself concretely. Did he mean say "to reify knowledge" instead of "ratify"? Perhaps he means that Internet postings certify (or verify?) knowledge based on data? Referring to computer representations, we might say, "*The internet's purpose is to verify models and theories through the accumulation and manipulation of ever expanding data.*" But it seems that he meant "ratify knowledge" to be something undesirable. And I remain unclear what is meant by the "the internet's purpose…"— a specific purpose? *Whose purpose?*

The sentences that follow are what most concern me: "Human cognition loses its personal character." How does this follow? How does using the internet as a tool cause someone's cognition to "lose its personal character"?

He then says, "Individuals turn into data, and data become regnant." The very essence of statistical studies (e.g., medical research) involves abstracting a population of individuals to create a database. Indeed, our privacy concerns go quite the other way, to be sure that the data are not identifiable as referring to particular people. And how does data per se exercise rule or authority ("become regnant")?

Notice that the graphics illustrating this article are intended to incite fear.



"This" technology referred to in the first sentence is purely imaginary—it not only goes "far beyond automation as we have known it," it goes beyond AlphaGo. Kissinger has extrapolated from a particular programming method that wins AlphaGo to an *"AI"* that does not exist. It's difficult to track the ensuing argument – "AI, by contrast, deals with ends; it establishes its own objectives." Assuming by "AI" he means "AI as we know it," in what sense is AlphaGo establishing its *own objectives*? It operates within the fixed ontology of a closed-world game—its designed objective is to win the game whose moves are all well-defined and fixed; it learns tactics, sequences of moves within a completely known world, not general strategies.

The question about consciousness is also scientifically imprecise. *Did he mean to ask, "What will become of human responsibility and agency…?"* The entire mention of consciousness here is off track. (Did he mean "conscience"?)

Does this imagined "AI" have *explanatory power* or not? Explanatory power to whom, for what purpose? Is the explanation "its own," that is internal and not conveyable to us (using a machine-learned language of terms and relations we cannot understand)? Would it be like the explanation of a physicist explaining quantum mechanics to a child? To the contrary, might artificial cognitive systems capable of conceptualization (which I emphasize do not exist) create and justify (causally, mathematically, historically) new concepts, which we might appreciate? These are big questions, but all vague and ungrounded in the reality of what AI technology can do, what we know about the conceptualization and consciousness, and by the actual methods we use to construct practical tools. **For a related critique of Kissinger's article see:** https://www.skynettoday.com/briefs/kissinger-ai

For contrast, see the positive perspective suggested by John Seely Brown— https://www.prgs.edu/events/commencement/2018/keynote-address-john-seely-brown.html.

[Title note: Cassius, a Roman nobleman, uttered this phrase when he was talking to his friend, **Brutus**, in **Shakespeare's** play Julius Caesar. The phrase goes, "The **fault**, **dear Brutus, is not in our stars** / **But in ourselves**, that we are underlings." (Julius Caesar, Act I, Scene III, L. 140-141).]

Unscientific, Cliché-Filled Analyses

Dystopian AI scenarios (Just-so science fiction stories about our future vs. scientific predictions) bear a resemblance to Intelligent Design (Just-so creationism stories about our past vs. scientific theories). Both illustrate mere storytelling and opinions, speculation versus scientific understanding based on evidence.

Notice the clichés: "Developing at breakneck speed"; "stay in the game." *Vague metaphors:* "Upgrade minds." *Pretentious pseudo-scientific terminology:* "human"

"Foreign words and expressions such as *deus ex machina* … are used to give an air of culture and elegance." *George Orwell, 1946*

"If you simplify your English, you are freed from the worst follies of orthodoxy. You cannot speak any of the necessary dialects, and when you make a stupid remark its stupidity will be obvious, even to yourself. Political language—and with variations this is true of all political parties, from Conservatives to Anarchists—is designed to make lies sound truthful and murder respectable. and to give an appearance of solidity to pure wind." *Orwell 1946*

Given that our cognitive capability to reflect on experience, develop causal stories, to anticipate future events, and thus to theorize and plan is fundamentally based on our consciousness, a process of categorizing what we have seen and done, and our own thoughts, how could intelligence "decouple" from consciousness? Harari also says, "Armies and corporations cannot function without intelligent agents, but they don't need consciousness and subjective experience."

(Challenge of AI lies in ourselves, continued)

Technology-Centered Design

Using current technology appropriately — to help people do what they want to do better — and improving it, requires understanding how the technology complements people's knowledge and practices.

Example — Consider using the **Apple Maps** app to navigate while driving. It requires touching the screen — where is voice commanding? Why can't you interrogate the system? Why can't I ask, "Are we taking Interstate 280?" Why do I have to poke on a tiny screen to find the list of steps? The design does not relate to the fact that I am driving nor that I want to understand and potentially negotiate the route.

When the system suggests a revised route I have to read the screen and push a button. I accept the change and 10 minutes later the detour is over, we're back on the original path driving up 85 to Sunnyvale. I want to ask, "How did that work out for you? How much time did we save?" I want a performance evaluation so next time when I get such advice I can know whether to trust it or not.

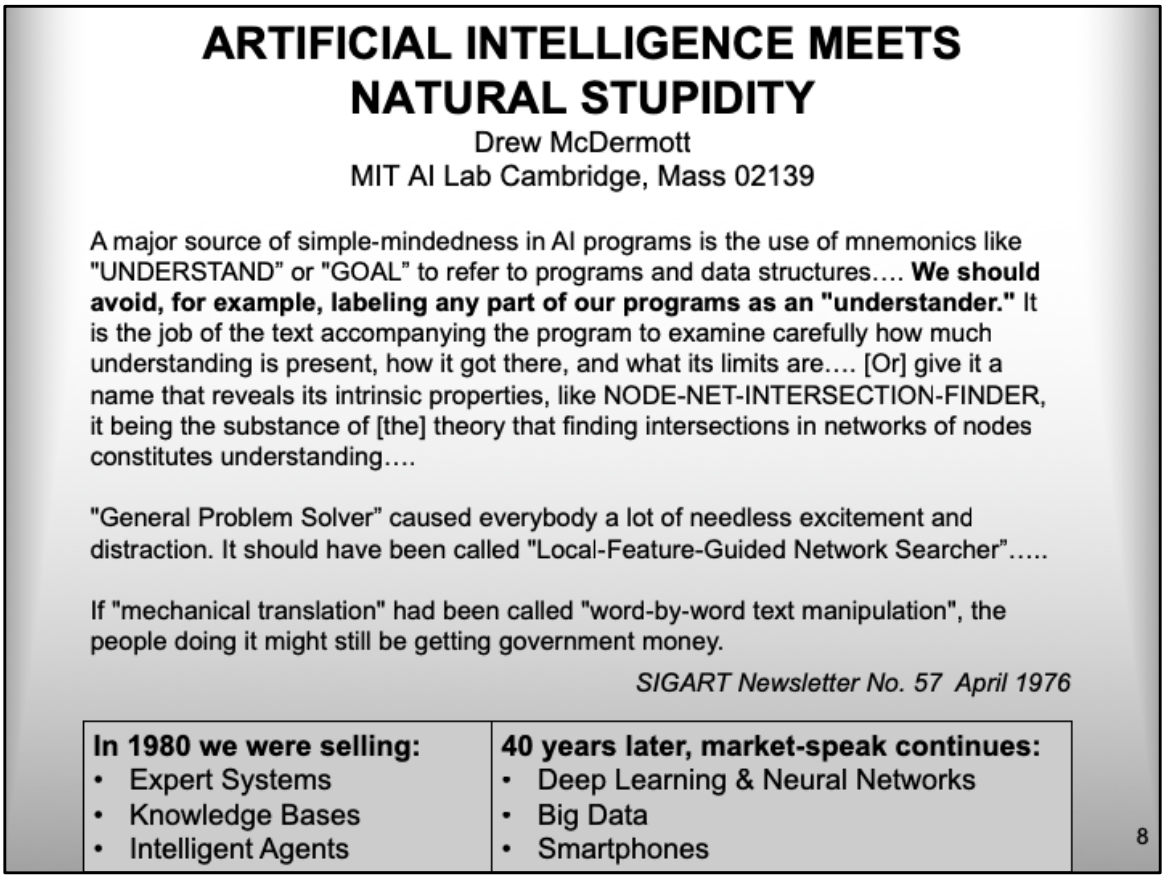
These shortcomings reflect organizational/social problems at Apple not technological — we built this kind of voice-commanded system at NASA 15 years ago (see the discussion of Mobile Agents which follows).

DARPA is putting $80 million into the Explainable AI program — how much is Apple with $100+ billion in the bank investing? These gigantic corporations are not leading the way. but millions of people rely on their software everyday — what's going to turn that around?

A key point here is complementing and augmenting our capabilities is very different from trying to replace people or merely automate what we do — that's a design principle Winograd and Flores emphasized 30 years ago in "*Understanding Computers & Cognition: A New Foundation for Design.*"

Nefarious Activities

The challenge of developing and applying AI wisely lies not in AI overlords, but in ourselves. Besides unscientific cliché-filled writing, hype, and systems that don't fit how we think and interact…**we have nefarious actors using AI and the social media for profit and power.** Kissinger's *Atlantic* article June 2018, "How the Enlightenment Ends," doesn't mention what is happening today and how current technology might be further abused to disrupt our infrastructure (e.g., electric grids, voting tallies, financial records) and manipulate opinion. Given actual, demonstrated harm from software already available, it is absurd to focus on technology we don't remotely know how to create.



This 1976 article in the ACM's newsletter for the Special Interest Group on Artificial Intelligence affected me very much when I was an AI graduate student in computer science at Stanford.

Notice how this unscientific use of terminology continues today. The 1980s systems were not experts; the program rules/schemas/frames/propositions etc. were models, not "knowledge": and the model-based agents were not a new breed of intellect.

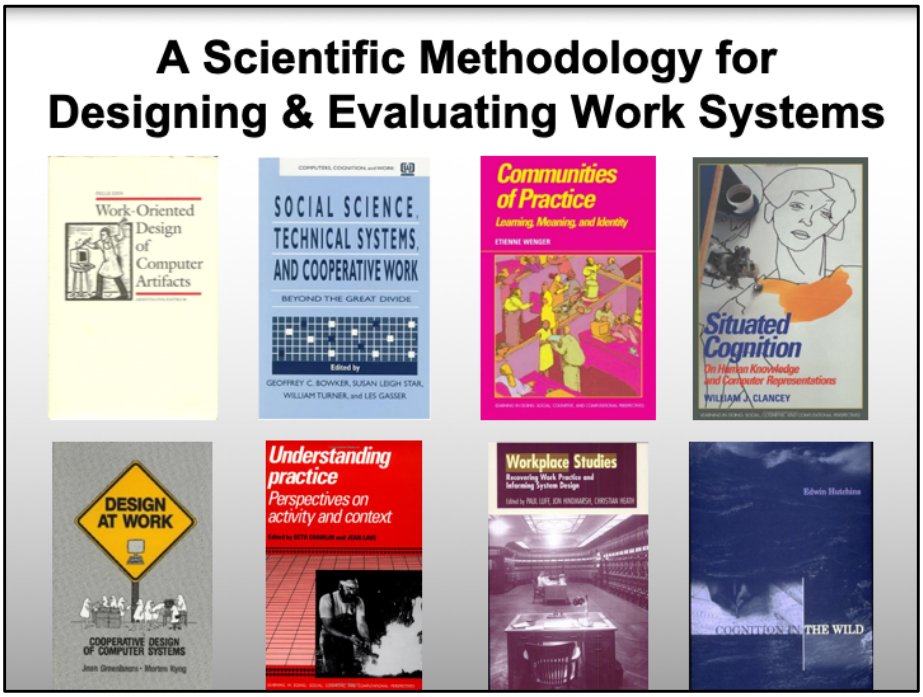
Today learning is "deep" and most researchers developing multi-layered recurrent architectures have no scientific commitment to modeling neural networks at all.

Part of the difficulty is that AI like other areas of engineering is not a single community of like-minded researchers, but consists of people who develop technical methods and those who develop practical tools These orientations are traditionally called “theory” and “applications.” But both are developing theory—one is grounded in math and logic, the other is empirical and scientifically oriented to people and the environment. Design methodology itself becomes a theoretical enterprise.

The book on the left is an excellent textbook. It should have been called AI PROGRAMMING TECHNIQUES. It doesn’t speak for all of “Modern AI” as the subtitle claims. The bibliography has about 2000 references; but the AI community that constructs practical tools is not represented at all. **Modern AI is both a set of programming techniques and a methodology for developing systems that fit how people think and behave (their practices)—**systems that are reliable, safe, and trustworthy.

The book on the right is a classic in human factors (1955). Dreyfus’s design firm was renowned for developing everyday devices (e.g. telephones, sewing machines, typewriters), equipment (e.g., the yellow-green John Deere tractor), etc. The book’s title is a fundamental theme of “Modern AI”—we design systems for people, which is to say, people are in control; the systems complement and augment what people want to do intellectually. (Eliminating physical labor or rule-driven office work is indeed a social problem, quite real and very different from effects of technology that “establishes its own objectives,” which Kissinger writes about.)

We need to highlight in our publications, talks, and interviews these parallel tracks of methods and theory. We need to emphasize that a large part of the AI community focuses on developing systems that are useful and fit people’s thinking and activities. Considering social implications must be part of this enterprise, too.



Over the past 25 years a work systems design methodology has been formulated and used in workplaces throughout the world including many corporations. The social perspective on work was pioneered in the 1980s by researchers from Scandinavia: Ehn, Bradley, Engeström. The research spread to the University of California in San Diego, Xerox-PARC, EURO-PARC, IBM, the Open University in the UK, and many other places.

We start by studying work practices so we can jointly formulate and understand the problem we are trying to solve and how a new work system design will affect the existing roles and responsibilities, activities, tools, layout, work schedule, work flow, etc. This design approach in Scandinavia was shaped by the social requirement for workers to be part of business automation; unions played a pivotal role. A US example: "One company that is realizing the benefits of union participation in work redesign is Corning….in partnership with the American Flint Glass Workers union (AFGW)…rooting out old production lines and retraining virtually all of its 20,000 employees to work in the new systems. About two-thirds of these workers need remedial education in reading and math." See : https://hbr.org/1991/05/what-should-unions-do *It is intriguing to consider how AI systems could enable people with only a high school education to participate in specialized, intellectual work.*

For a discussion comparing Dreyfuss's 1950s human factors approach to "total systems design" today, see the introduction of *Creative Engineering:*

Creative Engineering: Promoting Innovation by Thinking Differently. John E. Arnold (1959) 2016. Edited with an introduction and biographical essay by W. J. Clancey. Stanford Digital Repository: *http://purl.stanford.edu/jb100vs5745*.

In this section of the talk, I present systems we developed at NASA that illustrate a coherent work systems design methodology. These systems demonstrate how automation can enable people to do more of what they want to do, safely, more efficiently, and with higher quality.

The Mobile Agents project was inspired by an Apollo 17 event in which Gene Cernan asked Charlie Parker, the Capcom ¼ million miles away about the location of spare sample bags. Parker is virtually present with the astronauts via a television camera and radio. Gene didn't ask Harrison (Jack) Schmitt who was just a few meters away. How will we do this on Mars when conversations with Capcom in Houston are impractical because of a 5 – 20 minute one-way time delay?

This Apollo 17 video and transcript are available online.
https://www.hq.nasa.gov/alsj/a17/a17.sta5.html?fbclid=IwAR20msUg5kEwCsciXpJZvpiH084OJ57ocfOfLe9Nit791LEtAjthDn8I9Z8#1463019

The video on the right shows a demonstration of the Mobile Agents projects from 2005 at the Mars Desert Research Station in Utah (MDRS). Brent is on the left channel speaking to his personal agent who has a female voice; Abby is on the right speaking to her agent, who has a male voice. They are both geologists interested in this area, which they have never visited before, using the Mobile Agents system as a prototype tool to do authentic work.

See the MDRS 2005 video: https://www.youtube.com/watch?v=S4fjsIW2mk0

These are examples of capabilities demonstrated during the MDRS 2003 & 2004 Mobile Agents field experiments in Utah near the Mars Desert Research Station.

For details about the design methodology and system capabilities see:

Automating CapCom Using Mobile Agents and Robotic Assistants. *NASA Technical Publication, 2007.*

https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20070035904_2007036018.pdf

This slide explains the Mobile Agents Architecture. This is how we create a work flow system connected by agents. The request, "Boudreaux, take a picture of Astronaut 2" involves pointing Boudreaux's camera at Astronaut 2, which requires determining relative GPS locations.

The components–voice system, robot, camera, GPS–are made to appear as agents in the Brahms programming language by wrapping their APIs in "communication agents" (Cas), so these devices and Brahms agents can effectively communicate in a structured form of natural language (objects, attributes/relations, & values). CAs are Java programs that translate the API language of data structures and functions to the language of the task domain (e.g., "X take-picture-of Y"; "X follow Y").

Six agents are involved in processing this particular voice command. Astro-1 (Brent) speech agent → Astro-1 Personal Agent → Rover Agent → Astro-2 Personal Agent (what is Astro-2's location?) → Astro-2 GPS Agent. On receiving the location, the Rover Agent communicates to the rover's Camera Agent the command to take a picture centered on Astro-2's location.

We called this *Agent-Based Systems Integration*. From 2001–2013 the NASA Ames Mobile Agents Project used this method to configure into workflow systems a wide variety of cameras, instruments, rovers and robotic vehicles, displays, databases, and other software**.**

For description and analysis of the Mobile Agents workflow system configurations see: *https://billclancey.name/SMCIT2011WClancey.pdf*

As I have stressed, "AI" as a discipline today is more than a body of programming techniques and algorithmic theory. A multidisciplinary community of computer scientists, psychologists, and social scientists have developed a design methodology for creating practical systems through iterative experiments with prototypes in authentic work settings.

Design of Mobile Agents was informed by the Apollo experience and our personal experiences in role-playing "being on Mars" at the Mars Desert Research Station in Utah (shown in the slide). Rather than just asking engineers to design products for Martians, at MDRS people role-play being on Mars (Sol IV, a reference to J.E. Arnold's *Arcturus IV* case study), using their imagination to carry out a simulation of what living and working on Mars would be like.

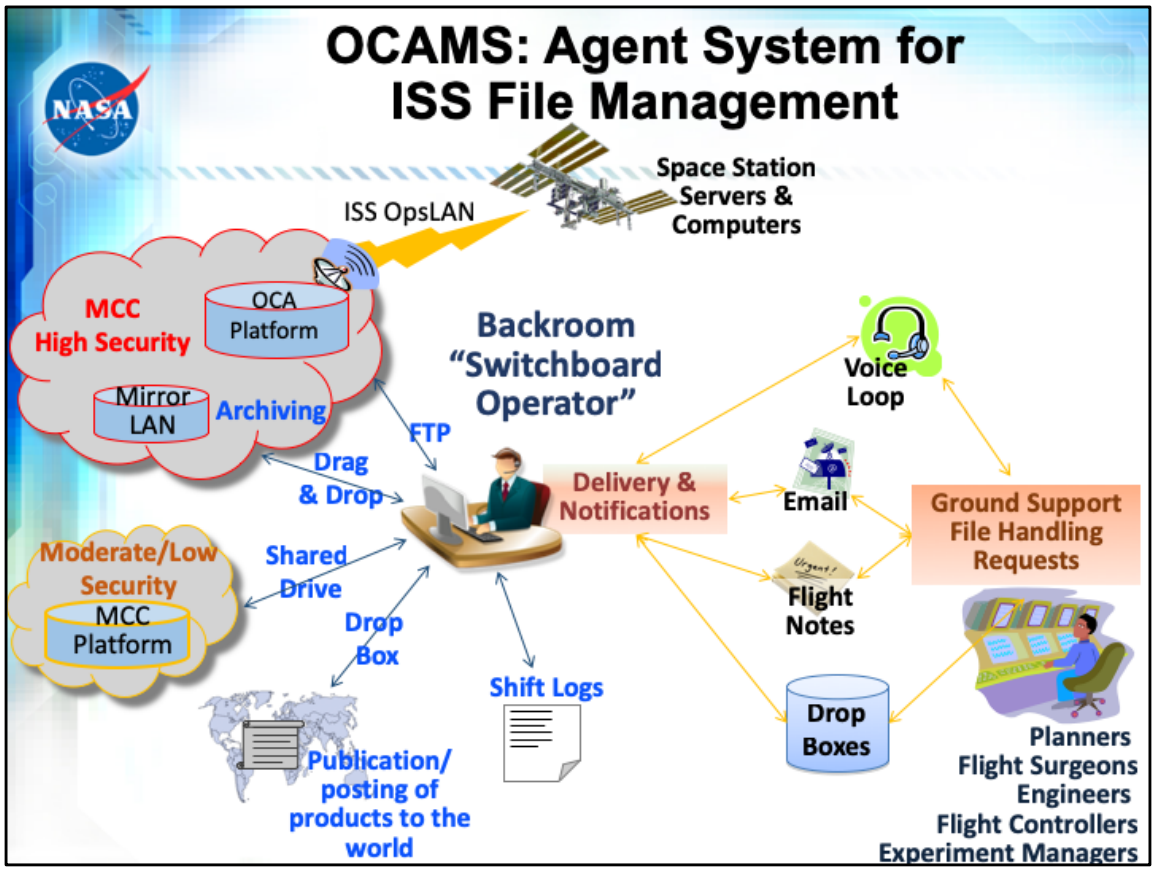
I've described MDRS in terms used by Douglas Thomas & John Seely Brown in *A New Culture of Learning: CULTIVATING THE IMAGINATION FOR A WORLD OF CONSTANT CHANGE.* They characterize innovation as involving a combination of play and imagination. That's not all that's required of course; you need technical methods (e.g., knowledge of AI programming), design processes, and tools. You need to engage in a disciplined R&D approach. As in world-building games described in "*A New Culture of Learning*," the simulated missions at MDRS provide a setting, experience, and learning environment that facilitates play and the imagination.

This is how we can develop "AI of the future" – not the technology that Harari and Kissinger describe, but interactive tools under people's control, which enables them to do what they want to do safely, reliably, and more productively.

Regarding our R&D approach—we developed the Mobile Agents system iteratively in an annual cycle of prototype experiments, designing improvements with the geologists based on their experience while using the system to do real work in a setting that interested them.

The engineers at NASA often do "requirements analysis" for imagined future missions. I called our methodology **Empirical Requirements Analysis** to stress that determining design requirements is a scientific problem and requires experimentation, with an ethnographic perspective of recording and reflecting on people's experiences.

By this methodology, R&D is a scientific endeavor. It is empirical, experimental, and is based on theories and models that relate potentially complex systems of people, technology, and the environment.

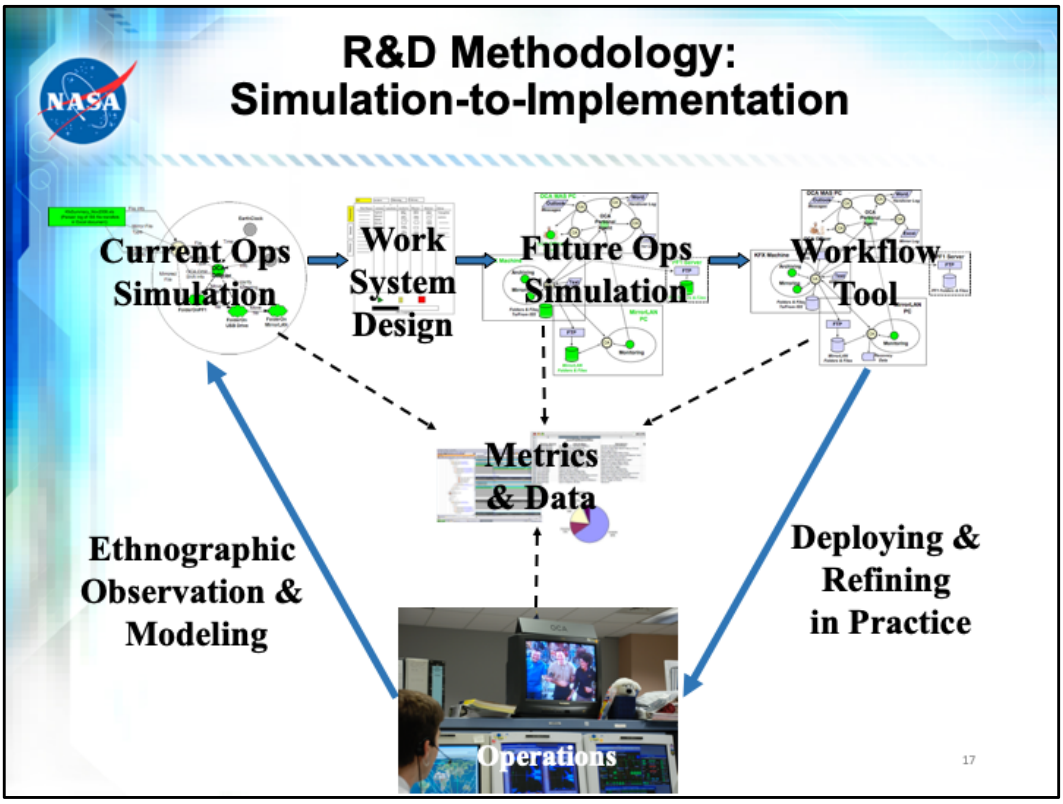


We also developed a program using the Mobile Agents architecture called OCAMS, which received the Johnson Space Center Exceptional Software Award, the first awarded to researchers at another NASA center. The problem was to remotely manage the file system onboard the ISS computers. The program consists of a collection of "agents" that integrate legacy hardware and software to create a single workflow system.

OCA = Orbital Communications Adapter – refers to a PC card in the computer used to transfer files from JSC backroom through Deep Space Network (TDRS satellites) to ISS computers. The OCA Officer was a backroom starter position; an aerospace engineer served as the "switchboard operator" between ISS and ground support.

In developing Mobile Agents for Mars explorers, we started with the idea that Earth support was not immediately available, so we sought to automate on Mars what Capcom on Earth was doing for Apollo astronauts. In the OCAMS system, we started by observing and collaborating with backroom ground support personnel to make their work more efficient. We automated the routine work of the OCA Officer, a kind of switchboard operator who supports ground support personnel by managing files onboard the ISS. Together we eliminated that 24/7 position, with a savings of millions over the lifetime of the International Space Station. The "payback" occurred (i.e., savings exceeded R&D costs) five years after the system was deployed. OCA Officers were reassigned to more creative and challenging roles.

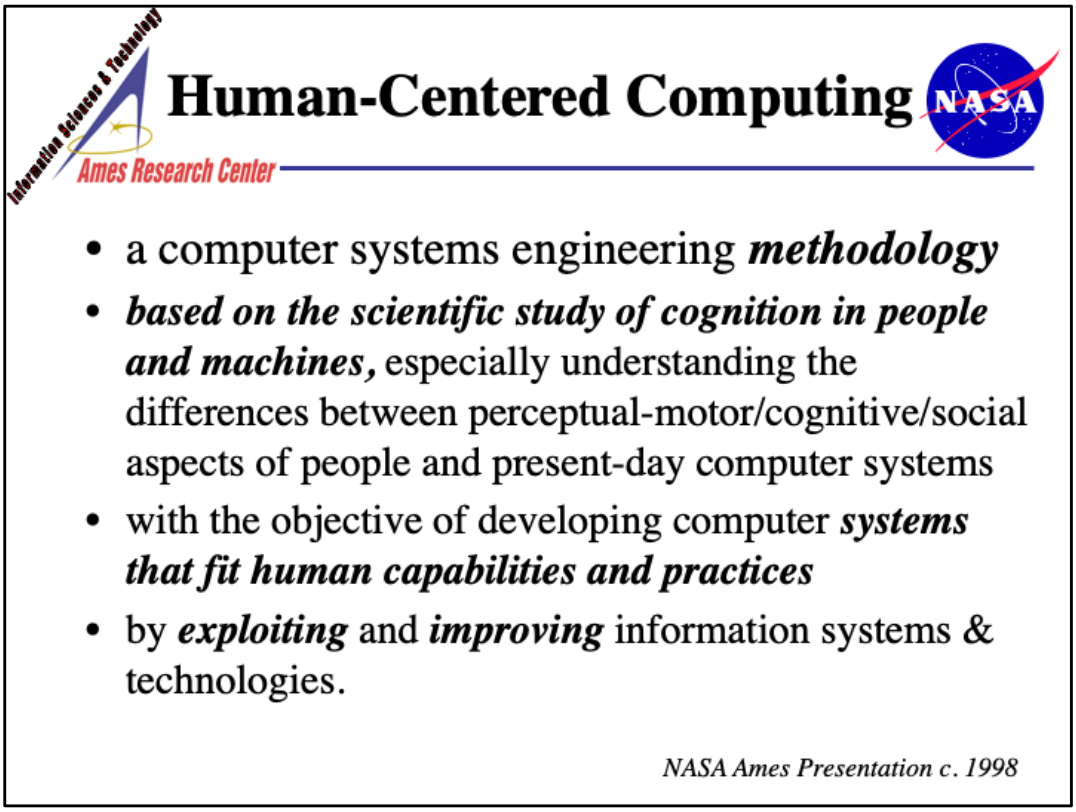
For details and further discussion see: Multi-agent simulation to implementation: A practical engineering methodology for designing space flight operations.



OCAMS was created using a *Simulation to Implementation* software development approach. It involves modeling and simulating current practices, developing a new work system design, and simulating it as the "Future Operations," all in the Brahms framework. The Future Ops simulation is created from the Current Operations model by "cutting" part of the OCA officer model and "pasting" it into a new agent, which became the OCAMS software program. Thus OCAMS automates some of the work previously done by the OCA Officer. In this new design the OCA Officer has a new activity of interacting with the OCAMS tool, whose interface is also modeled in the Future Ops simulation.

Simulation to Implementation is an example of how the method of design thinking can be combined with more traditional approaches in the field of engineering—observation, modeling, verification, and automation.

Using a Brahms activity-based simulation for design demonstrates how the framework enables capturing, augmenting, and adapting successful practices to create a new, more efficient and productive work system. Now the OCA Officer is freed to spend more time on difficult matters, handling other job responsibilities, and providing more customized one-off service to ground support personnel. Eventually the OCA role was combined with another backroom position.

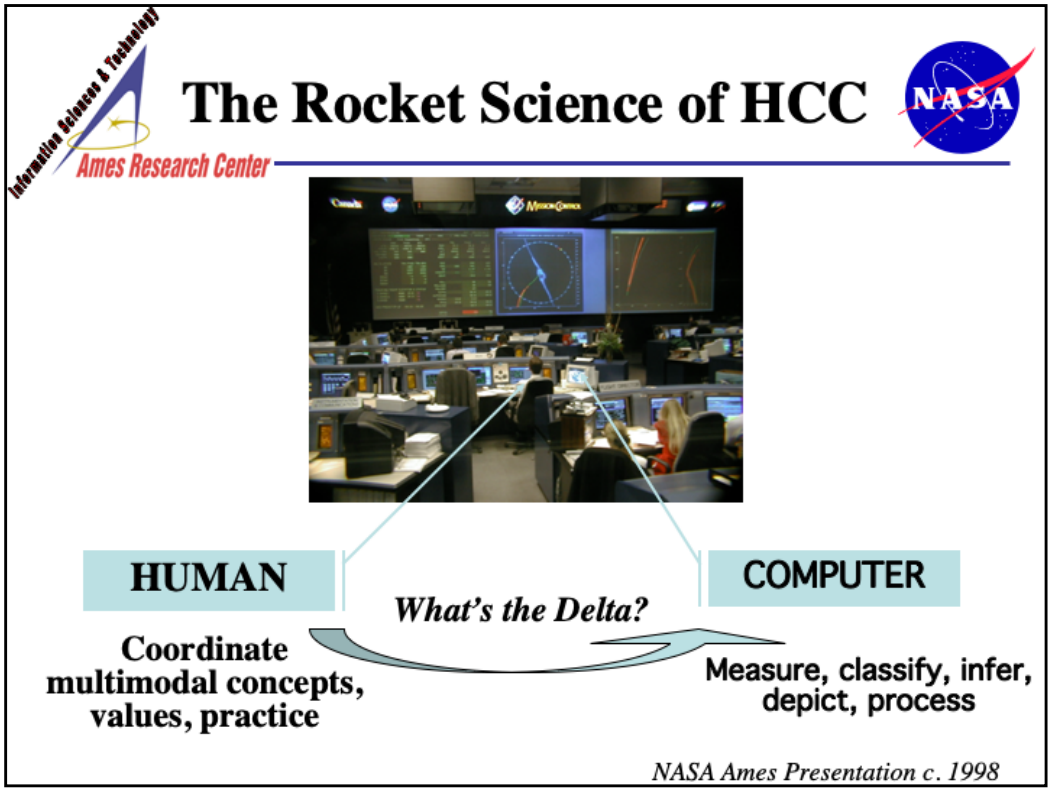


Stepping back a bit, I wanted to show you one of the first slides I developed at NASA when I was invited in 1998 to lead Ames research in an initiative they called "Human-Centered Computing." (More about this use of the term "human" later.)

I stressed from the start that HCC was a methodology for developing socio-technical systems.

< READ SLIDE >

Relating this back to the discussions about AI today, "Superintelligent beings" controlling human affairs are not going to suddenly materialize among us. That's not the kind of systems we are building or could build. Instead, we have well-established methods for developing tools that fit what people are trying to do, complementing and extending their capabilities. If we deploy systems that people cannot understand, and are thus out of control, that will show itself to be a problem and indeed catastrophes may occur—e.g., the Chernobyl nuclear plant meltdown. That's a fault of the designers, managers, shareholders of corporations, and citizens, not a fault of some artificial being who has taken over society. (Is imagined "boogey AI" a projection onto technology of our own out of control and unethical desires?)



When created in the late 1990s, HCC was defined as a core thrust in NASA Ame's Center of Excellence in IT. HCC blends the research of the cognitive, social, and computer sciences to both advance AI and to use existing technologies optimally.

The essential idea is this: If we don't understand the differences between people and the best computer systems today, we won't understand where people need help and how their tools might fail them. If we understand the relative strengths and limitations of people and computers, we can build synergistic, robust systems.

But also, AI stagnated in the 1980s because we lost track of our target, we started equating humans and automation. We talked about knowledge as if it could be written down exhaustively and stored. We started talking about judgement as if it were only a kind of decision-making calculus. We equated belief systems with sets of logical expressions (and called it "truth maintenance"), idealizing and objectifying belief, and we lost the subjective, cultural values of judgement and expertise.

One essential difference between people and computers is that people can coordinate ideas in different modalities--verbally, visually, aurally, in gestures. But no computer system has ever conceived of anything at all—all they can do is manipulate data descriptions, classify, make logical inferences, and model relationships. No machine that we know how to build can improvise and coordinate behaviors in the way people can.

We Design for People not "Humans"

When would it be appropriate to say, "I saw five humans standing in line at the grocery store"?

20

The term "HCC" was given to me at NASA, but I restricted the usage to my title as Chief Scientist, HCC, Intelligent Systems Division, NASA Ames. But in my own writing I usually only use the term "humans" to distinguish our primate species.

Consider this photo. Who would say "I saw five humans standing in line at the grocery store"? A psychologist? A zoologist? A science fiction character who is more familiar with robots picking up groceries?

Using the term "human" allows us to treat people as identical. But when a system design must reflect people's interests, what they are trying to accomplish, and how they want to work or interact with tools, then we must acknowledge we are dealing with individuals, that is, *persons*. To me "human-centered" means "abstracted, not personal, not individualized to a person's interests and ways of doing things."

Use of the term "human" makes me suspicious–to what extent is it a guise of objectivity, a way of appearing scientific? To what extent does "human-speak" impose a theoretical framework, a certain set of eyeglasses, on what we are studying? How would it change our seeing, our understanding of needs, and our vision of what could be in the future if we saw people as *persons?*

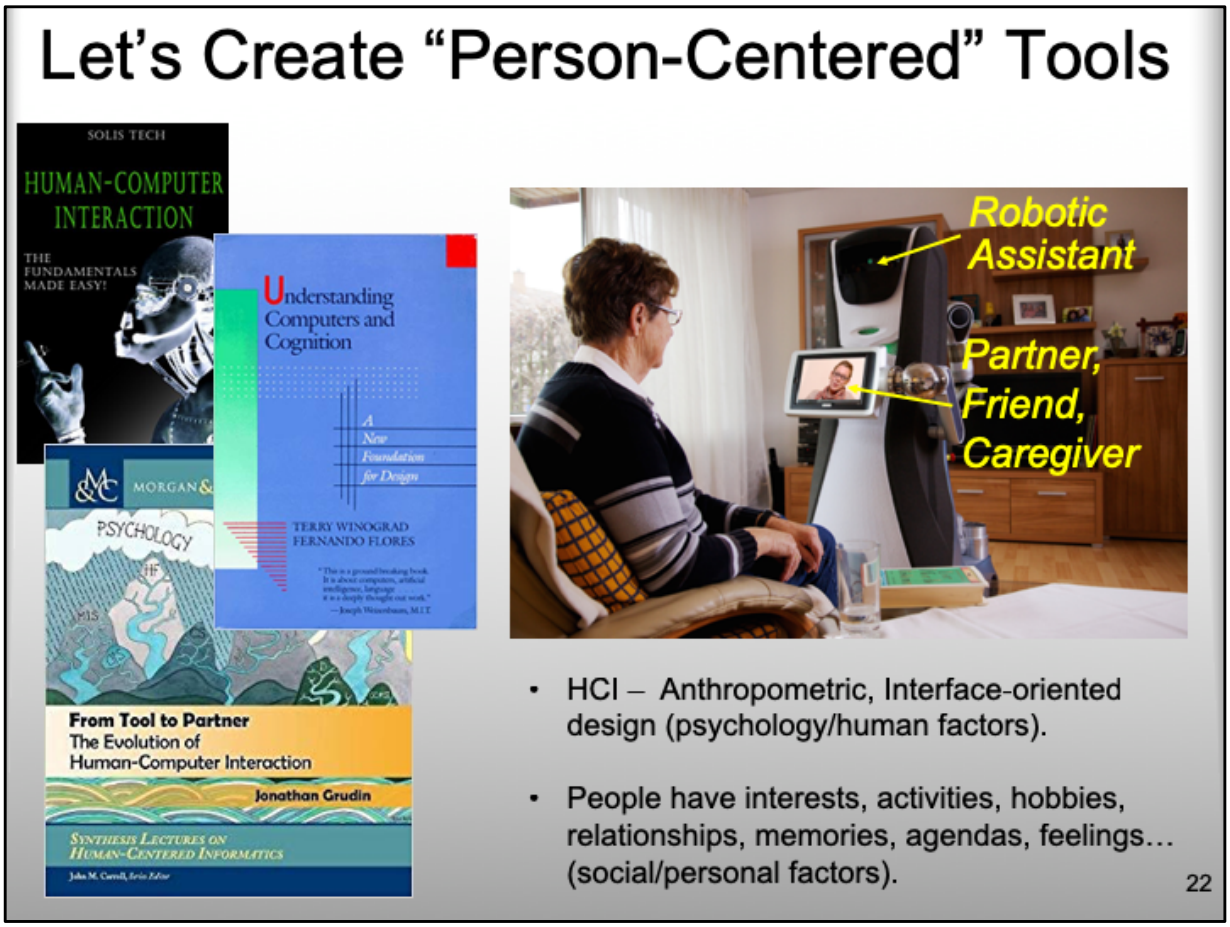
Side point: The photo illustrates what is meant by "practices –cultural behaviors, often tacitly learned through mimicking, ways of sustaining order in our interactions. In institutions practices are often deliberately constructed as written procedures and policies. How people stand in line (or not) in different countries demonstrates that practices are cultural. In a functional abstraction, we say people carry out the same *task* (e.g., buying groceries), but they do it differently; they have different practices.

Referring to people as people rather than “humans” is part of my crusade to speak clearly and more insightfully about what are doing and what we have accomplished.

One of the leaders relating clear speaking (and seeing) to good design is Don Norman. I found this excellent interview with him with Joe Posner on VOX.

He speaks plainly with common sense observations from which he makes radical suggestions—that we pay attention and figure out how to make devices whose use is discoverable and that provides feedback so we know that (and how) our action was effective. He mentions in this interview that at a certain point he realized “user-centered” sounded strange—why not call it “people centered”?

See the video online: https://99percentinvisible.org/article/norman-doors-dont-know-whether-push-pull-blame-design/



What would it mean for a tool to relate to me as a person? Imagine a person-centered robot assistant for the elderly. Following the design heuristic in Winograd & Flores' book, *Understanding computers and cognition,* we might *facilitate conversations* between people rather than automating conversations (replacing people). Computer-aided instruction might mean bringing students and teachers together on demand, when guidance, explanation, or evaluation is helpful—rather than (only) replacing the teacher by emulating "intelligent tutoring."

Furthermore, when we think in terms of *designing for a person*, we realize we might take into account –

- Emotions
- History, experiences
- Personal interests
- Personalities
- Intellectual proclivities

I expect that Grudin's survey book is worth reading, given his reputation and experience. But "From Tool to Partner" is an example of the inflation of terminology. Do we need every person who assists us to be a partner? Do I need the person who check out my groceries to be a shopping partner? When librarians assist our research, do we need them to collaborate on our projects?

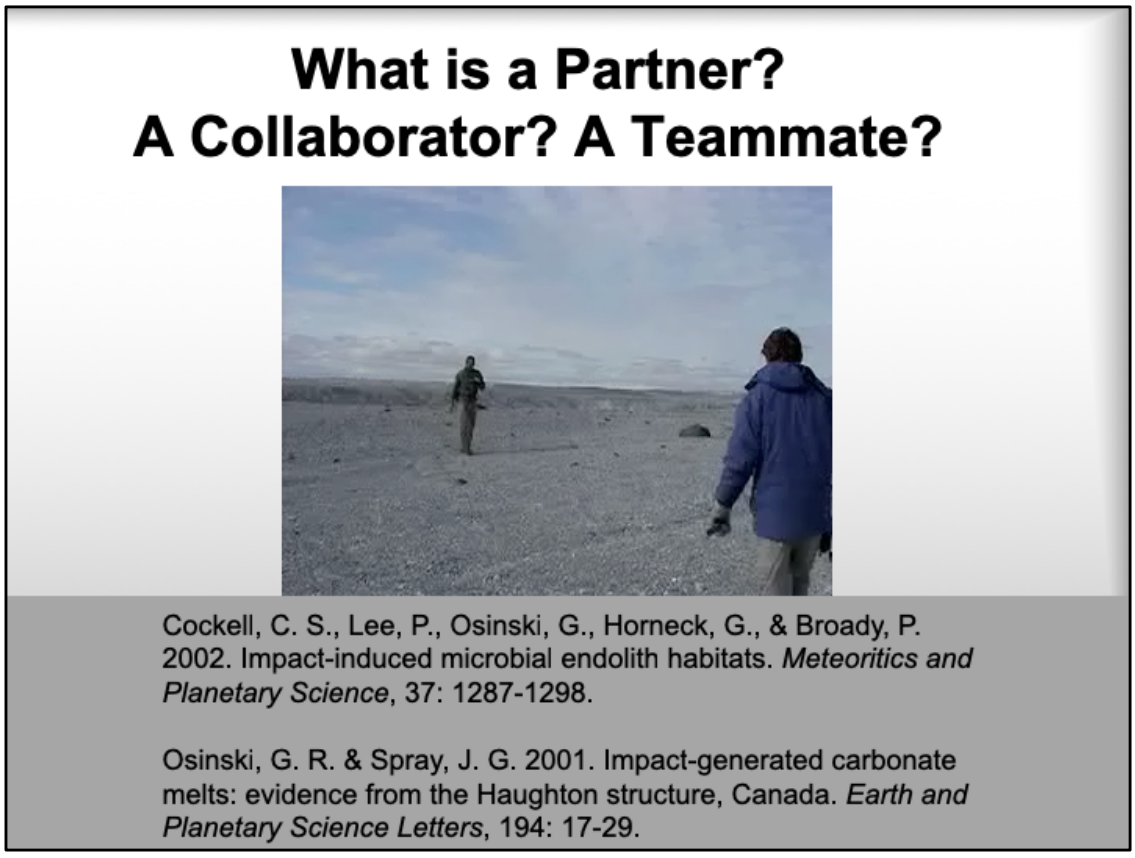
Here is an example of a personal project: I'm studying to get a ham radio license so I can serve as the neighborhood point of communication in a fire evacuation or an earthquake. Notice how this learning activity, this project, creates an identity by enabling me to participate in a certain way—I will have a new role in my community.

Do you have hobbies? Pursuits for self-improvement, skills you are developing such as playing a music instrument? Sports? A personal project requires having a sense of the self; that's a central effect of consciousness—conceptualizing "What I'm Doing Now" (WIDN) etc. This is ongoing tacit reflection, categorizing our activity and state of mind.

Antonio Damasio's neuroscience analysis reveals how emotion is not the opposite of reason; it's essential to reason. Emotions assign value to what we perceive and do. If you don't know what you want, you can't make good decisions. Caring and concern orient us to observe, to discern, and to consider what is and what might be.

Conceptualization is …

- Bound to emotional qualities and feelings, e.g., interest, value (liking, wanting, avoiding…), and needs (desire, hunger, fatigue…).
- Blending categories multimodally, across verbal, visual, auditory, olfactory, kinesthetic systems.
- (Re-)creates processes, which can be sequenced, substituted, and composed like things (the discrete "digital" aspect of habits, mental models, and speech).
- An aspect of remembering (see Bartlett's book by the same name); it enables constructing causal stories (of ourselves, our experiences, our partners, and our world).
- The basis of planning and designing (scheming) and of scientific modeling and theorizing.



During an afternoon traverse at Haughton Crater in 1999, I observed Osinski, a geologist, collaborating with Cockell, a biologist. But the biologist was not part of the geologist's project. Their collaboration was asymmetric. This distinction is reflected in the co-authorship of their reports about their fieldwork at that site.

If you want to develop a robotic partner or collaborator, that's fine. But to do that you must know what a partner or collaborator is. You must understand what it entails for someone to be your collaborator or teammate. In particular, you need to understand cognitively and socially what is required.

On what projects might a five year old child be your collaborator? Is everyone in your own specialty field capable of collaborating with you? Why not? You should consider if "a robotic partner" is what you really need. Would an assistant be more appropriate?

We must use such terms meaningfully, if they are provide guidance in designing new automated systems. And that always begins with empirical studies. What are your work practices? What would improve the quality of your work?

For further discussion and a transcript of the video in this slide (pp. 7–8), see:

Roles for agent assistants in field science: Understanding personal projects and collaboration. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 34, No. 2, May 2004. https://billclancey.name/IEEEClanceySMC.pdf

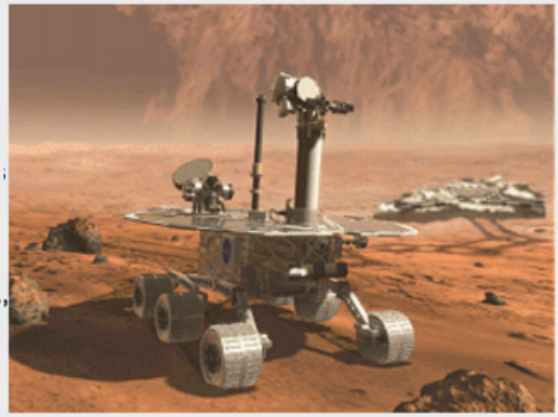
NASA Press Release Announcing MER Mission

Back to the Future on Mars

NASA announces plans for a Mars rover in 2003 with a second rover under consideration.

July 28, 2000 – In 2003, NASA plans to launch a relative of the now-famous 1997 Mars Pathfinder rover. Using drop, bounce, and role technology, this larger cousin is expected to reach the surface of the Red Planet in January, 2004 and begin the longest journey of scientific exploration ever taken across the surface of that alien world.

"This mission will give us the first ever robotic field geologist on Mars…"



MER is a mobile robotic laboratory, not a geologist.

If the only word that ever follows "intrepid" is "explorer" in print, then the words are choosing the meaning. *(after George Orwell, 1946)*

To appreciate the shift from loose poetic talk to scientific accuracy that I'm advocating, notice how the Mars Exploration Rover (MER) mission was first announced. MER was described as being a relative of Pathfinder and as being a field geologist. This press release is what originally motivated me to write *Working on Mars.* I knew that this simplistic anthropomorphizing didn't capture how the scientists would use the rover and so would fail to characterize what the automation was accomplishing—and that would hamper our thinking and might affect our funding for developing more sophisticated systems. And although the public might not be confused, they wouldn't understand the robotic technology either.

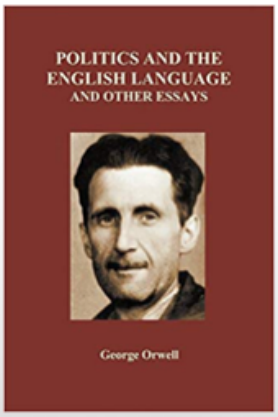
In explaining what the robots can do and why they are so important (and fascinating), I wouldn't speak this way to a five year. Why does NASA describe its technology this way? Is it too complicated to explain to the public the workings of a robotic laboratory? "Robotic geologist" is the language of marketeers and copywriters.

For an archived copy of the press release see:
https://science.nasa.gov/science-news/science-at-nasa/2000/ast28jul_2m

See also the article, "Human-Rover Partnership":
https://mars.jpl.nasa.gov/mer/fido/humanrover.html

The first line is: "Since humans cannot go to Mars yet, the Mars Exploration Rovers will act as robotic scientists." Would this be acceptable in a high school science class?

Orwell's Maxims



i. **Never use a metaphor, simile or other figure of speech which you are used to seeing in print.**
ii. Never use a long word where a short one will do.
iii. If it is possible to cut a word out, always cut it out.
iv. Never use the passive where you can use the active.
v. Never use a foreign phrase, a scientific word or a jargon word if you can think of an everyday English equivalent.
vi. Break any of these rules sooner than say anything barbarous.

"[The English language] becomes ugly and inaccurate because our thoughts are foolish, but the slovenliness of our language makes it easier for us to have foolish thoughts…. Modern English, especially written English, is full of bad habits which spread by imitation and which can be avoided if one is willing to take the necessary trouble. If one gets rid of these habits one can think more clearly, and to think clearly is a necessary first step towards political regeneration." (1946)

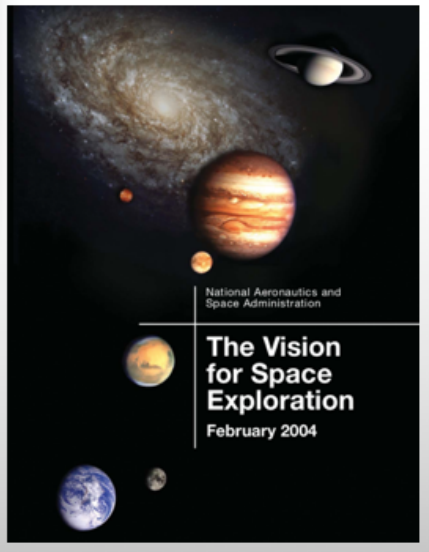
26

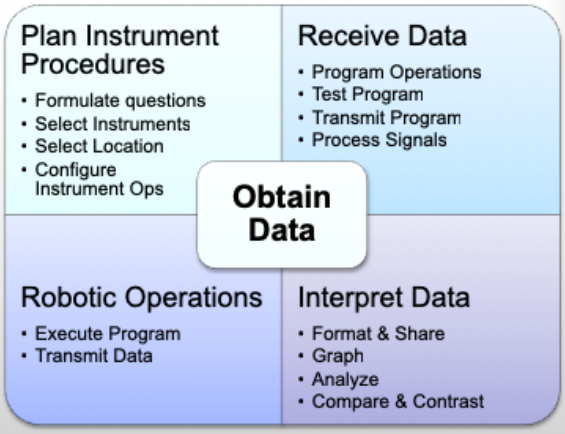
An example of "double-think" is the name **Department of Defense.** It was previously called the Department of War (Army & Navy & Air Force after 1947). It was renamed in 1949, the same year *1984* was published. Political "double-think" inhibits understanding what is really occurring. Orwell's 1946 essay is relevant to political rhetoric today:

"In our time, political speech and writing are largely the defence of the indefensible … Thus political language has to consist largely of euphemism, question-begging and sheer cloudy vagueness … the great enemy of clear language is insincerity. Where there is a gap between one's real and one's declared aims, one turns as it were instinctively to long words and exhausted idioms…."

In scientific presentations, the problem is not insincerity, but rather minds that aren't in gear, researchers just mimicking the jargon of others. In my experience this reflects a lack of analysis or one that is superficial and strives to establish an "authoritative voice." Renaming the "Heuristic Programming Project" at Stanford to "Knowledge Systems Laboratory" was a theoretical claim that equated program structures with the knowledge of experts; thus, the programs were "expert systems." Next, some researchers called their program's representations "deep" and characterized others as "shallow." In "Heuristic Classification" (*Artificial Intelligence*, 27:289-350, 1985) my analysis of AI programming "representational" frameworks demonstrated that the researchers had invented different ways of *modelling processes* (e.g., diseases, circuits; diagnosis, planning). A similar commitment to scientific language and terminology is required in AI research today.

Clear thinking requires clear speaking





"What is above all needed is to let the meaning choose the word, and not the other way about."

George Orwell, Politics & the English Language, 1946

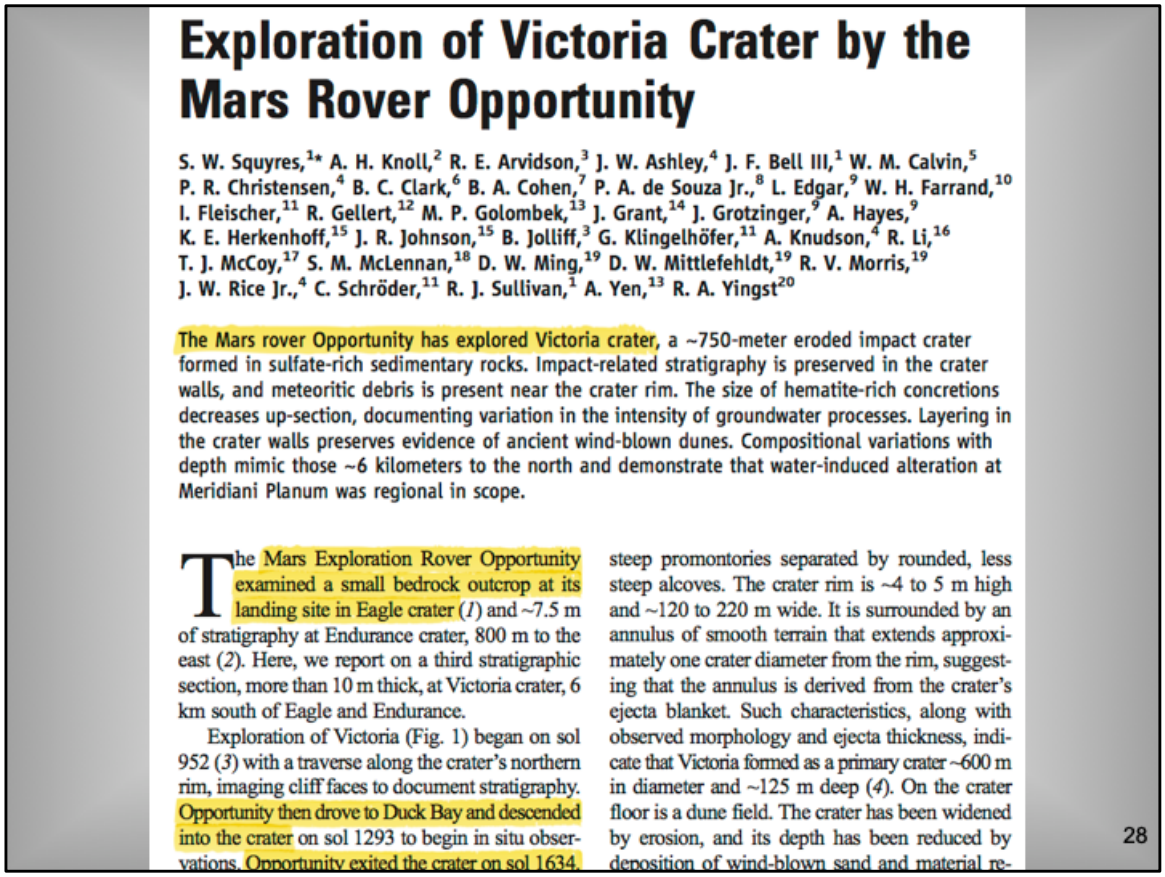
27

NASA's VISE 2004 mission statement says:
"NASA will send human and robotic explorers as partners, leveraging the capabilities of each where most useful. Robotic explorers will visit new worlds first, to obtain scientific data, assess risks to our astronauts, demonstrate breakthrough technologies, identify space resources, and send tantalizing imagery back to Earth."

Orwell would call "Robotic Explorer" a "prefabricated phrase." Notice the slovenly term "partner" and the cliché "leveraging capabilities." Who is obtaining scientific data, assessing risks, etc.? Who is selecting what photographs will be taken? Notice the shift from the subject of "visit new worlds" to the rest of the sentence.

This kind of talk is mesmerizing. But it doesn't characterize what we are trying to do, what we are accomplishing, and how to do it better. It leads to talking about how few engineers can control the robot or how many robots an engineer can control, rather than enabling more scientists to participate in the mission team.

I outline some of the activities required to "obtain scientific data." The robotic operations involve receiving and executing a program and transmitting data. NASA's sloppy description of planetary science in the 2004 "vision" and elsewhere—characterizing remotely controlled instruments as "robotic partners"—depersonalizes the space program, obfuscating people's work and motives. It requires a socio-technical analysis to explain how people use robotic systems and when having people on the Moon or Mars might be essential. Instead the copywriters provided a glib "vision," which not surprisingly was soon forgotten.



Exploration of Victoria Crater by the Mars Rover Opportunity

S. W. Squyres,[1*] A. H. Knoll,[2] R. E. Arvidson,[3] J. W. Ashley,[4] J. F. Bell III,[1] W. M. Calvin,[5] P. R. Christensen,[4] B. C. Clark,[6] B. A. Cohen,[7] P. A. de Souza Jr.,[8] L. Edgar,[9] W. H. Farrand,[10] I. Fleischer,[11] R. Gellert,[12] M. P. Golombek,[13] J. Grant,[14] J. Grotzinger,[9] A. Hayes,[9] K. E. Herkenhoff,[15] J. R. Johnson,[15] B. Jolliff,[3] G. Klingelhöfer,[11] A. Knudson,[4] R. Li,[16] T. J. McCoy,[17] S. M. McLennan,[18] D. W. Ming,[19] D. W. Mittlefehldt,[19] R. V. Morris,[19] J. W. Rice Jr.,[4] C. Schröder,[11] R. J. Sullivan,[1] A. Yen,[13] R. A. Yingst[20]

The Mars rover Opportunity has explored Victoria crater, a ~750-meter eroded impact crater formed in sulfate-rich sedimentary rocks. Impact-related stratigraphy is preserved in the crater walls, and meteoritic debris is present near the crater rim. The size of hematite-rich concretions decreases up-section, documenting variation in the intensity of groundwater processes. Layering in the crater walls preserves evidence of ancient wind-blown dunes. Compositional variations with depth mimic those ~6 kilometers to the north and demonstrate that water-induced alteration at Meridiani Planum was regional in scope.

The Mars Exploration Rover Opportunity examined a small bedrock outcrop at its landing site in Eagle crater (*1*) and ~7.5 m of stratigraphy at Endurance crater, 800 m to the east (*2*). Here, we report on a third stratigraphic section, more than 10 m thick, at Victoria crater, 6 km south of Eagle and Endurance.

Exploration of Victoria (Fig. 1) began on sol 952 (*3*) with a traverse along the crater's northern rim, imaging cliff faces to document stratigraphy. Opportunity then drove to Duck Bay and descended into the crater on sol 1293 to begin in situ observations. Opportunity exited the crater on sol 1634,

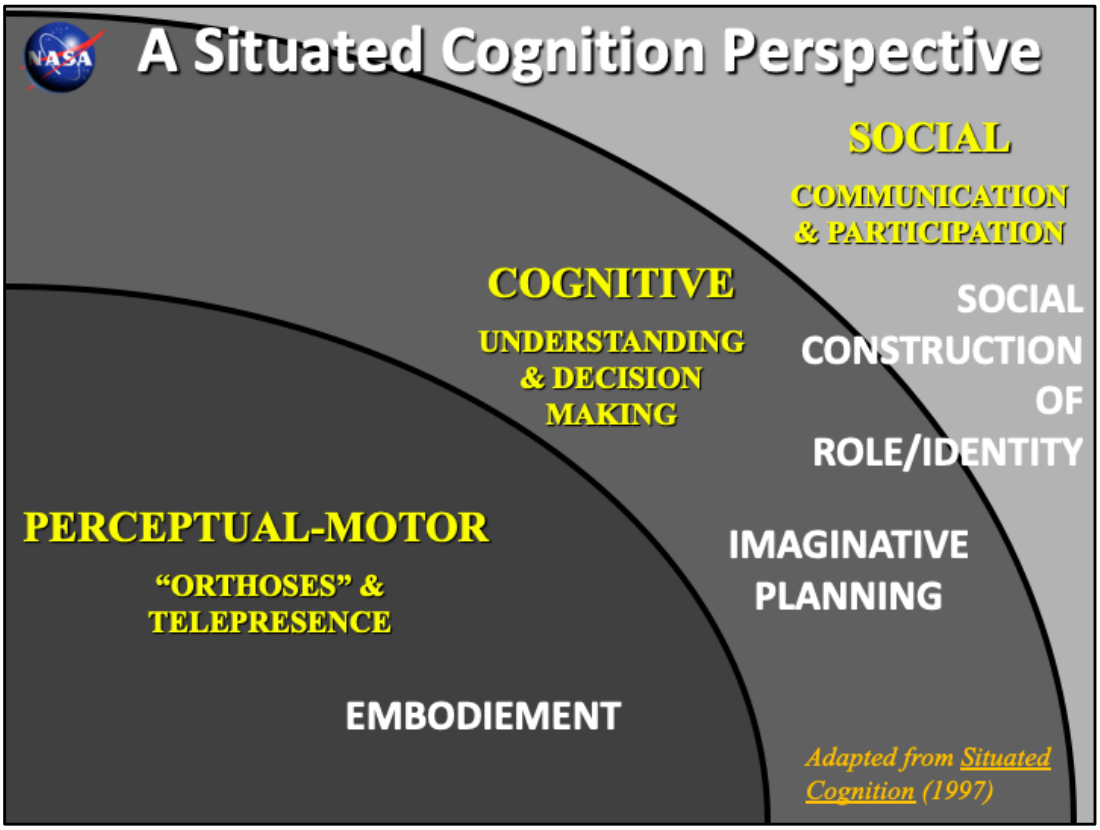
steep promontories separated by rounded, less steep alcoves. The crater rim is ~4 to 5 m high and ~120 to 220 m wide. It is surrounded by an annulus of smooth terrain that extends approximately one crater diameter from the rim, suggesting that the annulus is derived from the crater's ejecta blanket. Such characteristics, along with observed morphology and ejecta thickness, indicate that Victoria formed as a primary crater ~600 m in diameter and ~125 m deep (*4*). On the crater floor is a dune field. The crater has been widened by erosion, and its depth has been reduced by deposition of wind-blown sand and material re-

28

Besides continuing reference to the "robotic geologist" in the press, the scientists themselves describe the rover as if it was operating alone on Mars, exploring craters, examining rocks, and making discoveries. See: Squyres, Steve W., et al. "Exploration of Victoria Crater by the Mars Rover Opportunity." *Science* 324 (May 22, 2009): 1058–1061.

These are smart people, why do they write this way? What are they accomplishing? What values are they expressing? The discourse is third-person, projecting the team's actions into the rover. This style fits the team's consensus approach, acting as a group, and is an aspect of the social construction of "objectivity." The scientific culture requires reproducible abstractions, not particular to the experimenter or observer. Making the rover into the actor omits personal preferences and conflicts, which usually are irrelevant in a scientific article; it presents hypotheses, methods, data and interpretation, not personalities. See *Working on Mars* for discussion about the use of metonymy and its heuristic and social value for doing scientific work.

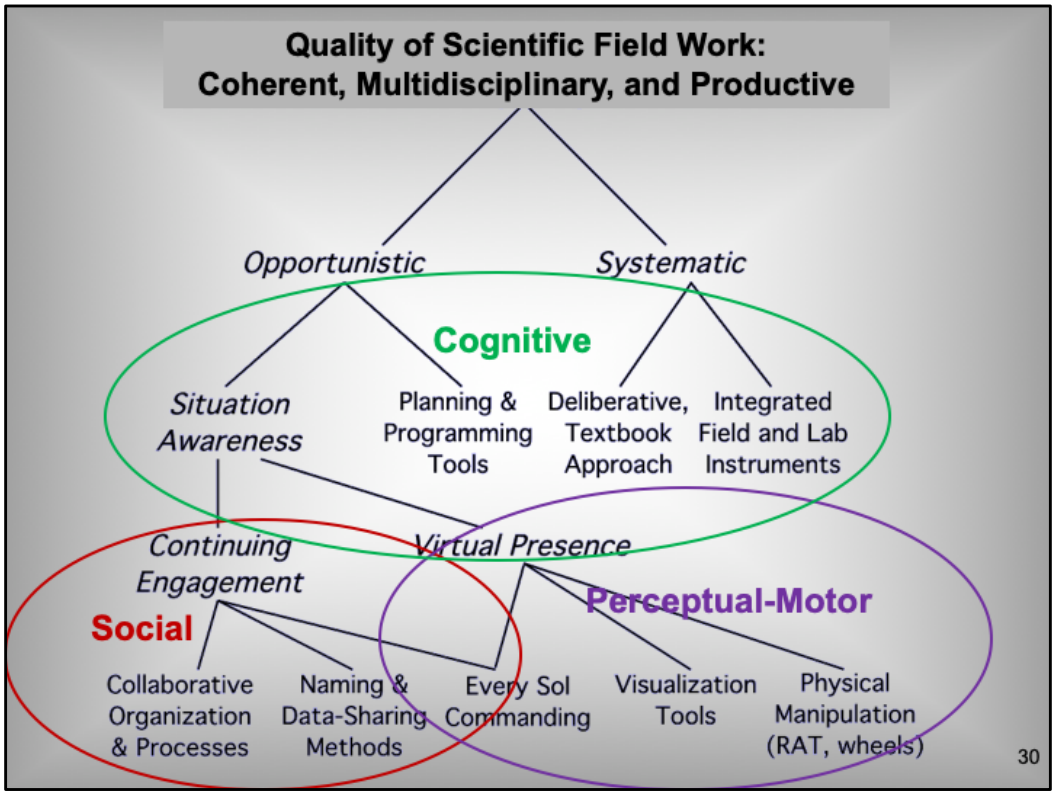
To build on MER's design NASA managers must appreciate why the work system was so successful. Disguising the role of people is detrimental for developing more capable tools or applying the socio-technical design to other domains. Similarly, a recent DoD report (DSB 2012) concluded that AI studies describing "levels of autonomy" have been "counter-productive [for design] because they focus too much attention on the computer" rather than on the joint work between the people and automated systems.



In analyzing MER, I was applying a scientific framework I first presented at NASA in 1998. The framework provides a basis for articulating the interacting activities and automated processes of a "socio-technical system" of people, technology, and facilities.

Each of the three analytic perspectives is important in the design the Mars rovers as an "exploration system"—comprising hardware, software tools; organization, roles, and scheduling of the scientists and engineers; and the facilities and layout of the work spaces.

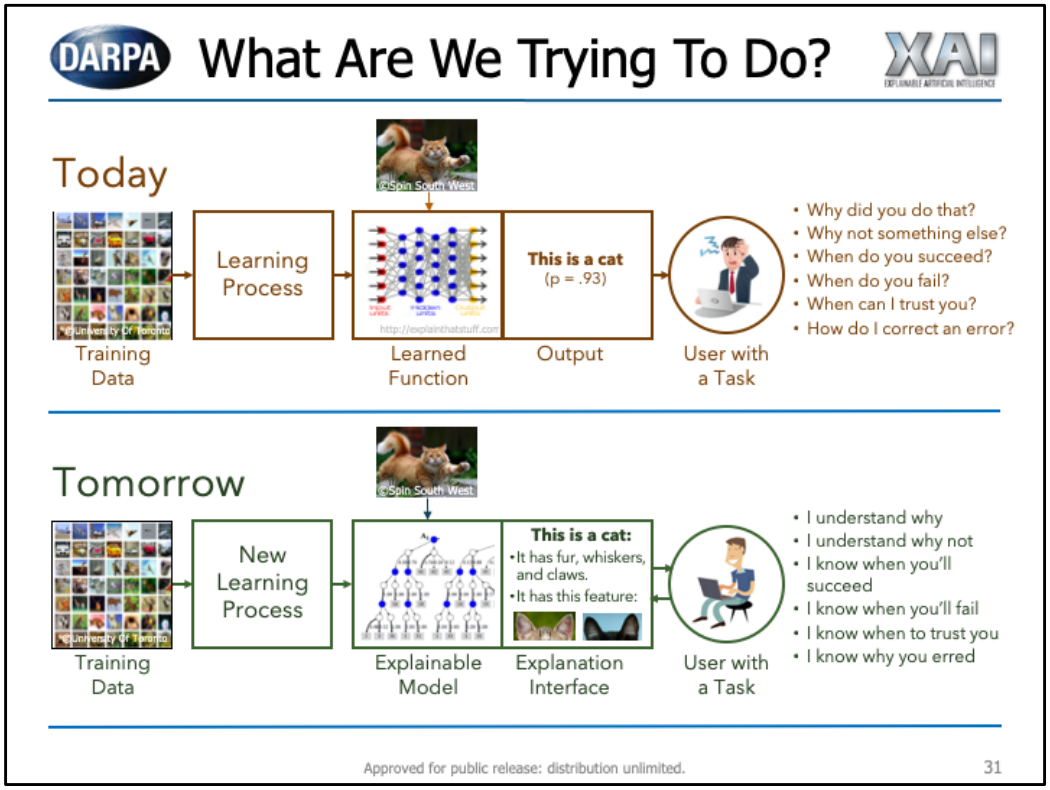
My study of MER led to refining this framework, which I earlier presented in another form in *Situated Cognition*. Conceptual frameworks develop over time from our readings, experiences, projects, and what we learn from other people. I am advocating that all researchers and journalists explicitly formulate their own understanding of the different aspects/dimensions of "intelligence." What eyeglasses do you use for describing and analyzing new technologies?



This diagram from *Working on Mars* sought to answer the question, "What accounts for the quality of the MER scientists' work?"

This kind of analysis is a counterpoint to the question a NASA Human Factors manager once suggested that I consider in my studies of work practices: "How can we mitigate the human?" Thus, some psychologists focus on people's limitations, failures, vulnerabilities (e.g., fatigue). That perspective has some value, but it says little about why people people do not fail most of the time. Social scientists at the Institute for Research on Learning taught me to start by asking, "What are the people doing well? What facilitates their success?"

This diagram and the framework on the previous page are *analytical perspectives* (accompanied by ways of observing and talking) by which we can understand the objectives, constraints, and resources the scientists integrated in constructing the MER exploration system. Accomplishing the *cognitive work* on Mars (scientific objectives, methods, and interests) involved *virtual perceptual-motor presence* (what they were perceiving and physically manipulating on Mars) and *continuing social engagement* over many years (personal and group activities). These aspects are mostly tacit, but when an impasse or conflict arises, issues relevant to quality are named and related in a discussion. For example, as the MER systems aged, some physical capabilities were lost, which rippled through the practices of commanding the rover, maintaining situation awareness, and following a deliberate scientific plan. The work process had to be re-coordinated physically and conceptually. See *Working on Mars* for a related discussion. (The diagram could be elaborated to show social and perceptual-motor aspects of "being systematic," such as rigorous scheduling.)



Slide from DARPA presentation, https://www.darpa.mil/program/explainable-artificial-intelligence: **"The Explainable AI (XAI) program aims to create a suite of machine learning techniques that:**

- Produce more explainable models, while maintaining a high level of learning performance (prediction accuracy); and
- Enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners."

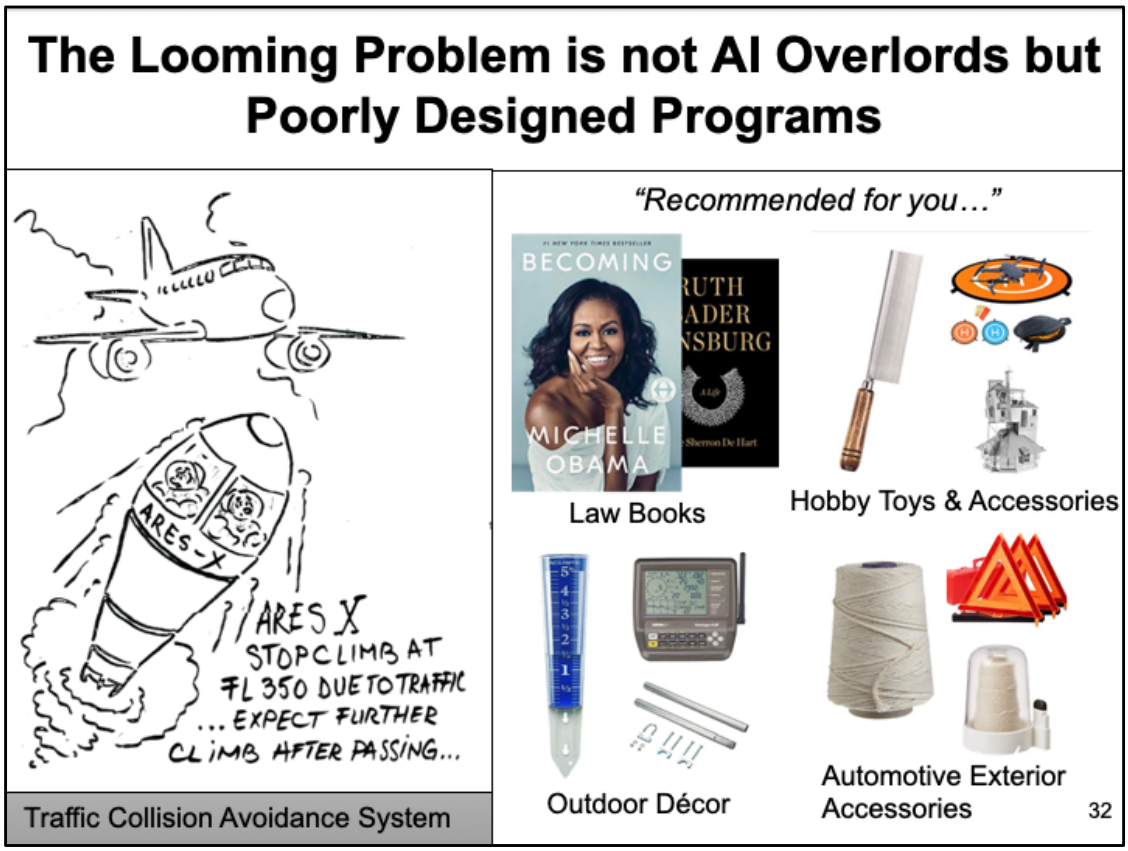
How might the vision for "Tomorrow" be broadened? What next steps might be advised for this research program? I see several issues:

1) "Explainable" does not entail "understandable, useful, and trustworthy." This is a technology-centered framing of the design challenge—the program can generate explanations.

2) Understanding and usefulness are context-dependent — does the person (aka "user," another technology-centered framing) have 10 seconds to make a decision, 10 minutes, 10 hours, etc.? Are other people involved or can other opinions be solicited? Is the system supplying data or advising actions?

3) "Explanations" are not usually things that are handed over, in the form of statements, diagrams, or other displays. "Explanations" often require an interaction, an activity involving follow-up, which may be incremental and iterative, rather than a linear "input → explanation → user" process.

3) The research is not going to produce "intelligent partners"—understanding how these tools relate to people's practices requires going far beyond this cliché.



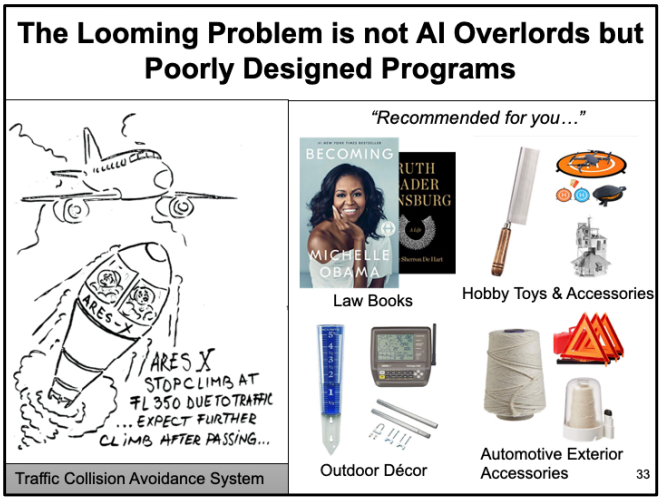
The faults in today's automated systems are caused by incompetent designers and managers, not out of control software. References to "the AI" of the future are distracting attention from actual shortcomings of today's designs and the organizations responsible.

On the left is a graphic from a publication about the Traffic Collision Avoidance system (TCAS) which is onboard all commercial flights. TCAS notifies pilots about traffic that may lead to a collision, and if the trajectories are not changed, tells the pilots what to do, usually to climb or descend. TCAS computers interact with each other behind the scenes to determine which way the planes should go, the Air Traffic Controllers (ATCs) are not involved. But what happens if the controllers discover the problem and tell the pilots to do something different? This led to a collision over Überlingen in 2002 that prompted a redesign of TCAS. Now if pilots disregard its advice, it tells the other plane's crew to reverse course. Is this a patch or a solution?

TCAS is an example of a system that tells people what to do, but was not designed as a total system that includes the ATC. It's a good example of fixating on a technical problem (preventing collision) while mostly ignoring the social-interactive aspect.

The graphic illustrates how TCAS lacks awareness of the context. Accordingly, operations policies state that it is advisory; the responsibility for safe flight remains with the pilots and air traffic controllers.

(continued)

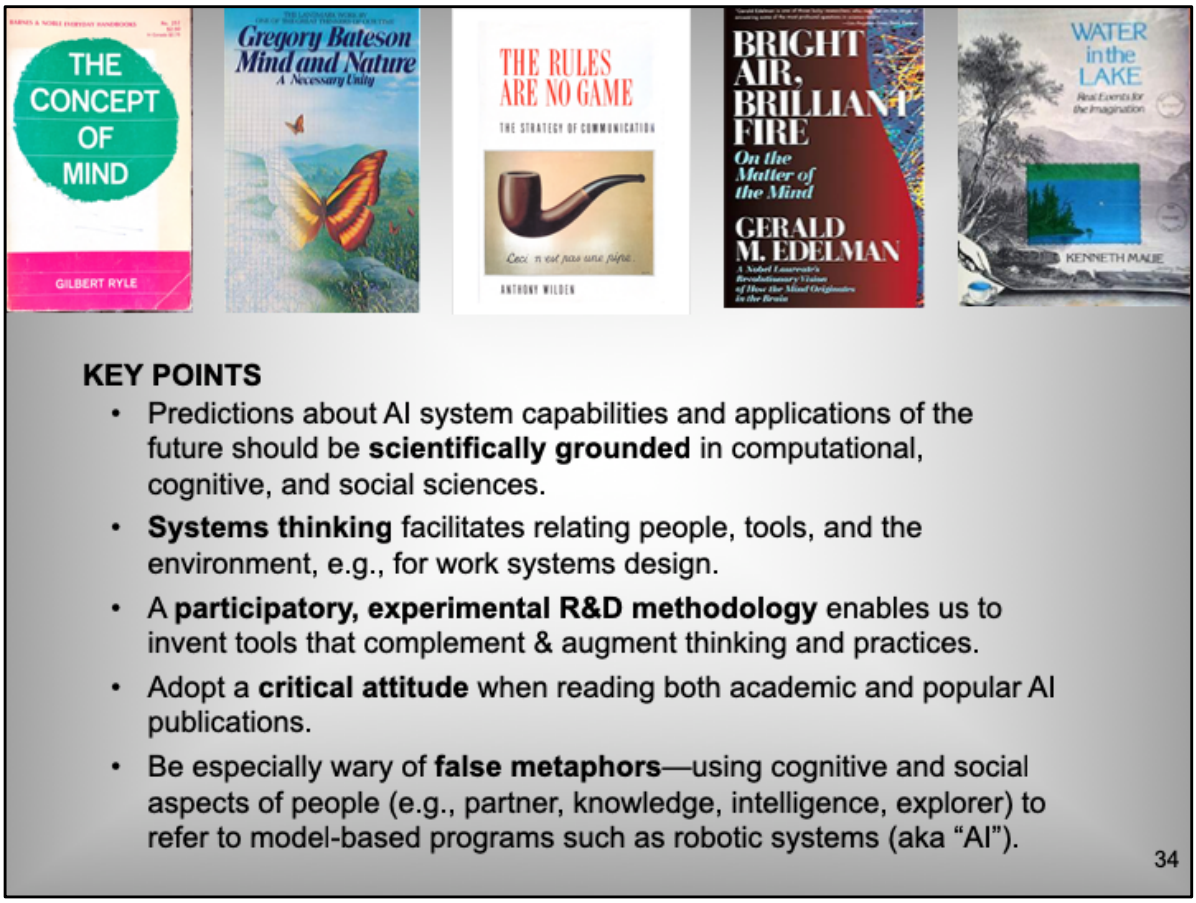


"Poorly designed programs" continued.

On the right are actual examples of items Amazon recently recommended for me to purchase. Notice how primitive the categorizing algorithm is. It's one thing to say Michelle Obama's book is about law, but simply wrong to refer to a weather station as décor or string as an automotive exterior accessory. And I wonder what hobby would involve such a large blade and wooden handle.

Where do these results come from? This is perhaps a real problem for the programmers who must figure out how to improve navigation programs like Apple Maps and online stores so their advice is correct, understandable, and we can trust them.

This is the design problem at hand, not AI overlords. Proper design requires systems that can explain their behavior, that can negotiate and receive feedback from us, and that relate to our experience, ways of working, and immediate interests. It ironic that with all of the emphasis on learning from "big data," far less attention is paid to learning from the customer.



Here are some books that have especially influenced me during my career. I recommend Edelman's book for the discussion of the neurological basis of consciousness and how it differs among animals. My book *Conceptual Coordination* builds on his theory, connecting it to cognitive models of language, planning, and reasoning. See the BBS article cited on the next page for a synopsis of my theory of conceptualization, illustrated by an analysis of dreaming.

My intent in this presentation has been to show that dystopian speculations about "future AI" and some of the concerns about machine learning are misguided. They are not based on the difference between computation and cognition in people. They use inappropriate metaphors and clichés. They are not scientific, but rash and often in the realm of science fiction. They do not reflect our best practices for developing automated systems. Nor do they reflect why systems like TCAS fail today. We must read and write clearly if we are to understand these issues scientifically.

We don't understand how conceptualization works well enough to build a machine than can replicate our capabilities—that entails simultaneously adapting, sequencing, and composing behaviors that are at once physical, reasoned, and emotional. Speculations about non-existent technology aside, what matters today is following well-established empirical design methods to develop tools that enable people to do better what they want to do. How AI is employed depends on the designers, managers, and people who use it to retain responsibility and to keep a critical mindset about what they are doing.

For more information…

- **Key References**
  - George Orwell, Politics and the English language. 1946
  - Joan Greenbaum & Morten Kyng. *Design at Work: Cooperative Design of Computer Systems*. 1991.
  - Oliver Sacks, *An Anthropologist on Mars.* 1995.
  - Murry Campbell, 2018. Mastering board games. *Science* 362: 1118.
- **Available at http://Bill.Clancey.name**
  - Viewing knowledge bases as qualitative models. *IEEE Expert: Intelligent Systems and Their Applications* 4(2), 9–15, 18–23, 1989.
  - Conceptual coordination bridges information processing and neurophysiology. *Behavioral and Brain Sciences*, special issue "Sleep and Dreaming," 26(3) 919-922, 2002.
  - Simulating activities. *Cognitive Systems Research* 3(3):471-499, 2002.
  - Roles for agent assistants in field science: Understanding personal projects and collaboration. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 34, No. 2, May 2004.
  - Scientific antecedents of situated cognition. *Cambridge Handbook of Situated Cognition*, pp. 11-34, 2008.
  - Simulating cognitive complexity in work systems. *Cognitive Science Annual Conference*, 2014.

See my web site, http://Bill.Clancey.name, for all of my publications and information about the book, *Working on Mars*. Contact – wclancey@ihmc.us