Martin Luther King and the "Ghost in the Machine"

Will Fitzgerald
Kalamazoo College
will.fitzgerald@kzoo.edu

Abstract

The U.S. Civil Rights movement and the field of Artificial Intelligence (AI) research had their beginnings in the mid-1950s, with no discernable interaction. The early AI researchers pursued a research programme that has led to a strong emphasis on rational action and cognition. As a result, AI researchers have lost opportunities to explore crucial aspects of agency, such as *forgiveness* and *justice*, the kind of themes explored by Martin Luther King. I argue that, AI researchers should take the opportunity to reexamine themes discussed by King and others both to explore these aspects of agency, and to create a more humane research programme.

To appear in *International Journal of Cognition and Technology* 2:1 (2003).

Martin Luther King and the "Ghost in the Machine"

Will Fitzgerald
Kalamazoo College
will.fitzgerald@kzoo.edu

Twins of different color

In 1955, John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude Shannon submitted "A proposal for the Dartmouth summer research project on Artificial Intelligence" (McCarthy, et al. 1955). This workshop, which was held in the summer of 1956 at Dartmouth College in Hanover, New Hampshire, was a "two month, ten man study" of "the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." Although there were certainly precursors to artificial intelligence research before this summer–Turing's paper on intelligence (Turing, 1950) and the field of cybernetics come immediately to mind–this workshop was the first to use the term "artificial intelligence" (AI).

Among the conference attendees were men crucial to the development of AI and computing in general. McCarthy, who was at Dartmouth at the time, went on to create the computer language Lisp, of fundamental importance to AI. Minsky and McCarthy founded the AI lab at MIT. Rochester, who worked for IBM, was influential in the design of early IBM computers. Shannon, of course, had already published "A mathematical theory of communication," (Shannon, 1948), the foundational document for information theory. Other attendees were Herbert Simon, who received the 1978 Nobel prize for Economics for his work on decision theory; Allen Newell, who, along with Simon had created the "Logic

Theorist," an early AI reasoning program that debuted at the conference; Arthur Samuel and Alex Berstein, also of IBM, who wrote influential programs that played checkers and chess; Ray Solomonoff, who did foundational work in machine learning and theories of induction; Oliver Selfridge, also a machine learning researcher; and Trenchard More, who developed array theory (fundamental to the programming language APL). McCorduck makes a strong case that the Dartmouth Conference was a "turning point" for the field, and that attendees of the first AI conference effectively "defined the establishment" of AI research, at least in the United States (McCorduck, 1979) for the next twenty years.

Also in 1955, Rosa Parks famously refused to give up her seat for White riders on a bus in Montgomery, Alabama. The bus driver called the police; they arrested Parks for breaking Montgomery's segregation laws which required her to give up her seat. The Montgomery Improvement Association was formed, and Dr. Martin Luther King was elected its president. The boycott lasted through almost a year, despite great financial hardship and violence against King and others, until the United States Supreme Court declared Montgomery's racial segregation laws unconstitutional. King achieved prominence as a result of the Montgomery bus boycott, and the struggle for civil rights for African-Americans became a national issue.

Both the modern U.S. civil rights movement and AI research were born in the mid-fifties. It will surprise no one that these two had little or no influence on each other. All of the attendees of the Dartmouth summer conference were White, were male, were from a small number of northern institutions–Princeton, MIT, CIT (later CMU), and IBM. The civil rights movement started in the U.S. South by African-Americans who sought basic human

rights: the right to vote, to use public transportation and other public accommodations freely, to work. Still, by taking no notice of the civil rights movement happening around them, the attendees at the Dartmouth conference may have missed some opportunities. What would have AI research been like, for example, if King had attended the Dartmouth conference?

King on high technology

King, of course, was not a technologist, nor did he write extensively about technology. When he did, he tended to be pessimistic about its goals and consequences. For example, he warned that automation might become a "Moloch, consuming jobs and (labor) contract gains (King, 1986b)." In his last Sunday morning sermon before his assassination, he said (King, 1986c):

> There can be no gainsaying of the fact that a great revolution is taking place in the world today. In a sense, it is a triple revolution: that is a technological revolution, with the impact of automation and cybernation; then there is a revolution in weaponry, with the emergence of atomic and nuclear weapons of warfare. Then there is a human rights revolution, with the freedom explosion that is taking place all over the world. ... Through our scientific and technological genius, we have made of this world a neighborhood and yet we have not had the ethical commitment to make of it a brotherhood.

Clearly, he hoped that high technology could aid the human rights revolution, but he feared it would not.

The first set of opportunities that AI researchers missed were to build a science that could be of peaceful service to the community of humanity. King saw the issue in stark terms: "In a day when Sputniks and Explorers are dashing through outer space and guided ballistic missles are carving highways of death through the stratosphere," he wrote in 1961 (King, 1986c), "no nation can win a war. The choice is no longer between violence and non-violence; it is either non-violence or non-existence. Unless we find some alternative to war, we will destroy ourselves by the misuse of our own instruments."

The ghost in the machine

The philosopher Gilbert Ryle spoke, with "deliberate abusiveness" of Cartesian dualism as "the ghost in the machine." The conjecture of the Dartmouth workshop ("every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it") is often described in dualistic terms: intelligence is just software that–so far–runs on the hardware of the human brain, but, in theory, can also run on computer hardware. This is often called the "Physical Symbol System Hypothesis" (Newell and Simon, 1976).

I do not wish to discuss whether the Physical Symbol System Hypothesis is *a priori* philosophically false. The expression "ghost in the machine" is very evocative, however, and many writers have used this image. In particular, I would like to quote from Gail B. Griffin's *Teaching Whiteness* (Griffin, 2003), a collection of critical essays, stories and reflections on being White and teaching at a largely White institution. Griffin picks up on the idea of the ghost in the machine to describe how Whiteness haunts:

> The irony (or paradox, or both) of whiteness is that its failure to name itself, while it arrogates one kind of godlike power (the power of universality and ubiquity), denies another. For to be universal and ubiquitous–to be Everything, Everywhere–is in fact to be Nothing, and Nowhere, in particular....As the absent agent in a passive construction, whiteness erases itself. White language says, in short, "I am not here; I do not exist." It does so, of course, to avoid implicating itself in the relations, past and present, of racism. But the price for such exoneration is eternal absence, non-being–ghostliness.

It is, perhaps, naive to think that the researchers who attended the 1956 Dartmouth workshop could have usefully viewed themselves as "White" (or even as male). But it is also, I think, significant that all of the original artificial intelligence researchers were White (and male), and this "ghostliness" of Whiteness hovers over AI research to the present. To pick up a copy of *AI Magazine* and look at the pictures of authors and participants is still to see mostly White, male faces. Again, this will come as a surprise to no one.

And yet this was another opportunity lost. It can at least be imagined that the early AI researchers could have, as the "active agent of an active construction," reflected on, written about, and modeled the effects of their own Whiteness on their science. A popular and excellent texbook on artificial intelligence (Russell and Norvig, 2003) states AI researchers tend to define AI as either a *descriptive* or *prescriptive* field of either *thinking* or *action*. The word used for *prescriptive* is *rational*. A *rational thinker* draws just the right conclusions from premises; a *rational actor* acts to maximize its best expected outcome. Does it go

without saying that defining intelligence in this way seems especially easy to do from a White and male perspective?

A naive model of idea discovery

As any AI researcher can tell you, doing research in artificial intelligence is hard. Everything seems to be tied up with everything else. It was thought that breaking the field into smaller components would be useful, and, of course, it is. But it turns out that to understand computer vision is just as hard as getting computers to use language as humans do; to program computers to create their own programs–to plan–is just as hard as drawing the right inferences at the right time. There is even an expression for this: "AI-hard." Language, vision, planning, and every useful subfield of AI seems AI-hard; and the field of AI can use all the help it can get to discover new ideas that push the field forward.

Here's a deliberately naive view of idea discovery: consider good ideas as being in a Platonic field, waiting to be discovered. Maybe they are in Erdös's "God's book of Proofs," maybe, like Kepler, we hope to think God's thoughts after him. Consider researchers as essentially experimental trials; each idea is likely to be discovered by a particular researcher with some probability. Assume, again, with deliberate naiveté, that the ideas are independent of one another, as are the researchers. Further, let's assume that each of the researchers is as good as each of the others, and that the ideas are all equally easy to discover. That is, we're assuming the probability–call it p–of any researcher discovering any particular idea is always the same.

Making these assumptions gives us a simple equation for how likely an idea is to be

discovered by at least one researcher: $1-(1-p)^n$, where n is the number of researchers. If the goal of AI is to discover with more good ideas, how can we do this?

Increasing the exponential

To discover more ideas, we want to increase $1-(1-p)^n$. One way is to increase p, but (given our naive model) this is the same for all ideas and researchers. The only other way to increase this number is to increase the exponential, n.

Perhaps someone has given you a time machine, and allowed you to hand pick the invitation list to the Dartmouth workshop, and you can invite up to 100 people, but just in proportion to their representation in the general population. In 1940, according to US census figures, about 43% of the population were White, non-Hispanic males–basically, the kind of people who were invited to the original workshop. If you just invite the 43, the probability of an idea being discovered is about 0.35[1] . If women are included (bringing the total to 88), this jumps to 0.59, and, with all US citizens represented, 0.64. If, in the present (based on 2000 census figures), you were given the same opportunity, just inviting White non-Hispanic males to the workshop is even grimmer, because the percentage has dropped to about 34%, so the probability is just 0.29. In other words, by including a full complement of people, we more than double the odds of an idea being discovered.

Colorless green ideas, sleeping furiously

We know that people are not colorless; it's likely that ideas are not colorless either. It's not

---

[1]This is with p=0.01.

surprising that the men of AI made great strides in discovering good ideas about rational action. It's also not surprising that the men of AI did not begin researching the importance of emotion on human thinking–the focus on "rationality" makes this a stretch. The intuition here is that the "color" of researchers and ideas effect the likelihood of the an idea being discovered.

Let's make the model just slightly less naive, and add "color" to ideas and researchers. Let's say that researchers are either blue or green, and ideas are either blue or green. Further, let's assume that it is much more likely for researchers to discover ideas of the same color than ideas of a different color[2] . Given 43 "green" researchers (representing the percentage of White, non-Hispanic men in 1940) and 0 "blue" researchers, and assuming that "green" and "blue" ideas are equally distributed, the probability of an idea being discovered is 0.21. With 43 "green" researchers and 57 "blue" researchers, this increases to 0.42. Using 2000 figures, 34 "green" researchers and 0 "blue" researchers yields only a 0.16 probability.

In other words, using 1940 figures, the odds of an idea being discovered using both "blue" and "green" researchers is about two times as great as using "green" researchers alone. Using 2000 figures, the odds are about 2.5 times as great. If we take a more "social constructivist" stance, modeling this by allowing the ratio of blue to green ideas to be equivalent to the ration of blue and green researchers, the odds of an idea being discovered

---

[2]For the examples below, I use 0.01 and 0.001, respectively. See Table 1 in the Appendix for details.

by blue and green researchers is about 3.5 times as great as using green researchers alone.

Dr. King is my research advisor

This model of scientific discovery is deliberately naive, as I said, and, as written can be criticized in two important ways. First, to the extent that the model is oversimplified, it may not be the case that increasing the diversity of a research team will yield new good ideas. Second, by using ”blue” and ”green,” it can be said that I am avoiding the real issues of race and gender for which they are proxies. And, to both of these criticisms, I will agree. Yet, I think the exercise is a useful one, for it brings out the question: are there research areas of what it means to be “intelligent” or ”human” that artificial intelligence should be exploring, but has not? Are there themes of “being human” or “being intelligent” that are *not* captured by the “rational agent” model?

Rereading some of King’s essays makes it clear that this is the case. Among the themes that King addresses are *justice*, *mercy*, *conversion*, *forgiveness*, *violence*, *revenge*, *race*, *politics*, *resistance*, *persuasion*, *honor*, *dignity*, *sacrifice*, *love*, and *evil*. And, of course, there are many more. A really good AI model of *forgiveness*, for example, is, I suspect, no harder to create than a good AI model of *temporal reasoning*, and no easier as well.

Now, to be fair, some researchers have investigated themes of this sort, especially by AI “scruffies” (who favor experimentation and reflection, in contrast to the AI “neats” who desire mathematical and logical formalization). For example, much of the research by Roger Schank and those associated with him (Schank, 1982; Schank and Abelson 1977) focused on story understanding, and it’s difficult to make progress in real story understanding

without focusing on themes such as these.

Still, as, Russell and Norvig state, "recent years have seen a revolution in both the content and the methodology of work in artificial intelligence. It is now more common to build on existing theories than to propose brand new ones, to base claims on rigorous theorems or hard experimental evidence rather than on intuition." Perhaps this means that the field is more mature; perhaps it just means the neats have won, and the rigor is rigor mortis, as Birnbaum claims (Birnbaum, 1991). But perhaps AI needs a renewal in the themes that it studies to open itself up to new, big ideas of what it means to think, to act, to be human.

What could AI be?

Martin Luther King had more important things to attend to than to participate in the 1956 Dartmouth workshop on artificial intelligence. Still, had he done so as a full participant, the field of artificial intelligence research might have followed different directions. AI missed several opportunities, but AI researchers can still pursue richer strands of research. I sometimes wonder whether, along with fields like "computational biochemistry" and "computational physics" and all of the other "computational *X*" fields, there couldn't be a "computational humanism" that could claim (and reclaim) some of the themes described by King and others, both building models of anything of what it means to be human, as well as being a model of an humane, anti-racist science.

References

Birnbaum, Larry (1991). Rigor mortis: A response to Nilsson's 'Logic and Artificial Intelligence'. *Artificial Intelligence*, 47: 57–77.

Griffin, Gail B. (2003). *Teaching Whiteness: The End of Innocence*. Unpublished manuscript.

King, Jr., Martin Luther (1986a). The American Dream. In Washington, J. M., editor, *A Testament of Hope: The Essential Writings of Martin Luther King, Jr.*, pages 208–220. Harper and Row, San Francisco.

King, Jr., Martin Luther (1986b). If the Negro wins, Labor wins. In Washington, J. M., editor, *A Testament of Hope: The Essential Writings of Martin Luther King, Jr.*, pages 201–207. Harper and Row, San Francisco.

King, Jr., Martin Luther (1986c). Remaining awake through a great revolution. In Washington, J. M., editor, *A Testament of Hope: The Essential Writings of Martin Luther King, Jr.*, pages 268–278. Harper and Row, San Francisco.

McCarthy, John, Minsky, Marvin, Rochester, Nathaniel, and Shannon, Claude E. (1955). A proposal for the Dartmouth summer research project on Artificial Intelligence. Technical report. http: //www-formal.stanford.edu/jmc/history/dartmouth.html.

McCorduck, Pamela (1979). *Machines Who Think: A Personal Inquiry into the History and Prospects of Artificial Intelligence*. W.H. Freeman and Company.

Newell, Alan and Simon, Herbert A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3):113--126.

Russell, Stuart J. and Norvig, Peter (2003). *Artificial Intelligence: A Modern Approach. Second Edition*. Prentice Hall Series in Artificial Intelligence. Pearson Education, Inc., Upper Saddle River, New Jersey.

Schank, Roger C. (1982). *Dynamic Memory: A Theory of Learning in Computers and People*. Cambridge University Press, Cambridge, UK.

Schank, Roger C. and Abelson, Robert P. (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Shannon, Claude E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27: 379–423,623–656.

Turing, Alan M. (1950). Computing machinery and intelligence. *Mind*, 59: 433–460.

Appendix: Idea discovery with color

There are Green ideas and Blue ideas. Let $P(G)$ be the proportion of Green ideas and $P(B)$ be the proportion of Blue ideas. There are green and blue researchers as well. Let $P(g)$ be the proportion of green researchers and $P(b)$ be the proportion of blue researchers. Researchers are more likely to discover ideas that are the same color as the researcher. Let $P(=)$ be the probability that a researcher discovers an idea of the same color and $P(\neq)$ the probability that a researcher discovers an idea of a different color. Then p, the probability of an idea being discovered, is:

$P(G)*P(g)*P(=)+P(B)*P(g)*P(\neq)+P(B)*P(g)*P(\neq)+P(B)*P(B)*P(=)$.

If N is the number of researchers, then $P(I)$, the probability of an idea being discovered, is $1-(1-p)^n$.

Year	*Researcher*	$P(G)$	$P(B)$	$P(g)$	$P(b)$	$P(=)$	$P(\neq)$	P	N	$P(I)$
Equal number of ideas										
1940	*White male*	0.50	0.50	1.00	0.00	0.01	0.001	0.0055	43	0.211
	White	0.50	0.50	0.49	0.51	0.01	0.001	0.0055	88	0.385
	All	0.50	0.50	0.43	0.57	0.01	0.001	0.0055	100	0.424
2000	*White male*	0.50	0.50	1.00	0.00	0.01	0.001	0.0055	34	0.171
	All	0.50	0.50	0.34	0.66	0.01	0.001	0.0055	100	0.424
Social constructivist										
1940	*White male*	0.43	0.57	1.00	0.00	0.01	0.001	0.0049	43	0.189
	White	0.43	0.57	0.49	0.51	0.01	0.001	0.0055	88	0.385
	All	0.43	0.57	0.43	0.57	0.01	0.001	0.0056	100	0.429
2000	*White male*	0.34	0.66	1.00	0.00	0.01	0.001	0.0041	34	0.129
	All	0.34	0.66	0.34	0.66	0.01	0.001	0.0060	100	0.450

Table 1: Naive model of idea discovery for ideas with "color". $P(I)$ is the probability of an idea being discovered $(1-(1-p)^n)$.

–