Attributing Intelligence to Humans and Machines: Between the Devil and the Deep Blue, See?

Carl W. Turner

Institute for Human and Machine Cognition
University of West Florida
Pensacola, FL 32504
cturner@ai.uwf.edu

Abstract

Human attributions of intelligence in others reflect heuristic judgments of superficial qualities rather than systematic analysis of competence. This article reviews the literature on attribution of intelligence in humans and machines, examines the public controversy over human intelligence, and extrapolates to the subject of machine intelligence. The possible effects on public policy are discussed.

Introduction

"By the end of 1978 computers existed which would crush 99.5% of the world's chess players....if this does not make you feel the computer's hot breath on the back of your neck, then nothing will" (Evans, 1979, p. 173).

The idea of computers successfully competing on an intellectual basis with humans arouses powerful emotions, as illustrated by the preceding quotation. Research on human reaction to computers usually deals with performance issues: does the machine adequately carry out its intended function, is it easy to use, is it cost effective? The outcome of the recent chess match between Deep Blue and chess champion Gary Kasparov, however, raised an entirely different set of issues in the minds of many. Deep Blue's chess rating places it among the top players in the world; its ability to play chess is unquestioned. In the days that followed its famous first meeting with Kasparov, though, most of the discussion was not over the details of this impressive technological accomplishment, but over a more subjective issue. The real question for many was, "Is the computer actually smart?"

The outcome of the historic first Kasparov-Deep Blue match was significant in at least one respect. Based on its widespread coverage in the news media, the match suggested to the general public that the advent of genuinely intelligent computers may be upon us. Theoretical issues aside, the match may have changed the way the public thinks about computers.

I propose to address the significance of the Kasparov-Deep Blue chess match in particular, and smart machines in general, in two ways. First, by reviewing the research on how humans attribute intelligence to both humans and computers. Second, by commenting on the cultural aspects of public attitudes towards human intelligence, and the implications for public policy. This article reviews the media coverage of and the public reaction to the current controversy over human intelligence, and extrapolates to the (coming) public controversy over machine intelligence. By doing so it may be possible to predict the trajectory of the coming debate, and to plan useful interventions.

Theories of Intelligence

Computer scientists who have skirted the issue until now are being forced to answer some difficult questions from interested lay persons. "Do you think the machine is really smart?" or the more thoughtful, "Do computers think like humans?" Often, they are unprepared to give a convincing answer. The stock response, "There is no agreed-upon definition of intelligence" is no longer satisfactory to a public that watched the Kasparov-Deep Blue chess match and has heard the reaction from both sides. True, there is no single agreed-upon definition of human intelligence, but there are a small number of competing classes of theories; Sternberg (1990) presents a good overview of the significant body of literature on human intelligence. The point of this article is not to review this literature, but to comment on another, equally intriguing phenomenon: the process by which humans subjectively perceive intelligence in others.

Attributions of Human Intelligence

Studies on the subjective ratings of intelligence are nearly as old as the original measure of intelligence, the IQ test. Cook (1984) provides numerous pointers to the research done from 1920's through the 1950's. Due to the methodological problems with subjective ratings and the IQ test, work in this area nearly ended in the 1950's. AI researchers with visions of defining the necessary and sufficient conditions for sensing intelligence in machines would do well to review some of the controversies over the subjective estimation of IQ.

After the 1950's, researchers sought to find the factors, both in the perceiver and the perceived, that influenced subjective attributions of intelligence and ability. Discussion of the factors that were found to affect subjective ratings of intelligence fills volumes: physical attractiveness, gender, age, physical handicap, and so on. The most interesting findings, and perhaps disturbing, were the apparently arbitrary test factors that influenced subjects' judgments of others' intelligence.

In one classic experiment, Jones et al. (1968) asked observers to rate the intelligence of other test participants (actually the experimenter's confederates) as the confederates took a phony verbally administered IQ test; the experimenter gave verbal feedback for correctness after every question. All of the confederates gave the same number of correct answers, but observers attributed higher intelligence to the confederates who answered more questions correctly at the beginning of the test than those who answered more questions correctly at the end. The reasons for this apparently arbitrary "primacy effect" on the attribution of ability are still unclear.

It is no surprise that statements made by experts on a particular subject are granted more credibility than those made by novices. However, the simple expedient of arbitrarily labeling statements as having been made "by an expert" is enough to increase a statement's believability (Petty, Cacioppo, & Goldman, 1981). The developers of expert systems should consider this finding before marketing their product as anything less than an expert.

"Everyone is a critic," as the saying goes, and for good reason; critics are often perceived as more intelligent than those who praise. Amabile (1983) presented subjects with edited excerpts of negative and positive book reviews. The negative reviewers were rated more intelligent than the positive reviewers, even when positive reviews were judged of higher quality.

The picture that emerges is that people hold implicit theories about intelligence on which they act. These implicit theories correlate to some extent with scientists' theories; intelligence includes several components, such as a problem solving component, a verbal component, and social competence (Sternberg, Conway, Ketron, & Bernstein, 1981). However, perceptions can be manipulated quite easily in the absence of defining data. Judgments of complex stimuli are made based on heuristics rather than on any systematic analysis (see Zimbardo & Leippe, 1991, for a review). Furthermore, introspection is not very helpful in discovering the factors that contribute to these judgments; much of this analysis goes on unconsciously. Thus, there seem to be no objective set of conditions for sensing intelligence, but a number of factors—some rather unintuitive—have been found to influence an observer's perception of intelligence in others.

Attributions of Machine Intelligence

Of course, the psychological research on human intelligence and attribution would mean little if the processes by which we attribute intelligence to machines bore no relation to the way in which we attribute intelligence to humans. In fact, there is evidence to suggest that humans may indeed sense intelligence in machines. One first-hand observer of the Kasparov-Deep Blue chess match was quoted as saying of the computer, "As it goes deeper and deeper, it displays elements of strategic understanding....This is the closest thing I've seen to computer intelligence. It's a weird form of intelligence, the beginning of intelligence. But you can feel it. You can smell it" (Weber, 1996, p. A1). This was not the romantic musing of a young science fiction fan, but the opinion of Frederick Friedel, Kasparov's technology-savvy advisor on computer chess. Clearly, there is a phenomenon here worth studying.

Recent work by Reeves and Nass (1996) attempted to demonstrate experimentally the idea that people behave towards computers as if the latter were social entities. The authors present a variety of clever experiments to test the hypothesis that electronic media, such as computers and televisions, elicit a social response in humans that was previously reserved for other humans. The goal of their experiments was simple: to find an interesting factor that affects the attribution of human ability and adapt it for a study of machines.

For example, to replicate the finding that praise and criticism have a differential effect on the perception of intelligence of a target individual (Amabile, 1983), Reeves and Nass ran the following experiment. Subjects were presented with lessons on a computerized tutoring system. After a brief test on the material presented, subjects moved to a different computer in order to receive feedback on the test. In addition to test feedback, the subjects' computers either praised or criticized the tutoring system. In follow-up evaluations, subjects rated the "critical" computers as significantly more intelligent than those which offered praise. Reeves and Nass also found effects of gender on the perception of a machine's competence (by supplying the computer with male or female voices) and of arbitrarily labeling one computer versus another an "expert" (Petty et al., 1981).

In arguments against the possibility of intelligent machines, much is made of the supposedly crucial "hardware" differences between humans and computers. The authors cite the argument made by Searle (1980) as a prime example. In fact, such an argument is irrelevant because of our human tendency to attribute intelligence to others, both human and human-like, based on whether they *appear* to think well. Reeves and Nass (1996) were not, of course, the first to notice that computers elicit strong social responses from people, but they are notable for having attempted to study the phenomenon from an experimental social science perspective, in an impartial manner, and without a crippling emotional response of their own (cf. Weizenbaum, 1976).

The debates over whether people believe that computers can think are not merely academic questions; the AI community should take note of the current public controversy over heritability and human intelligence.

Taking Measure of Intelligence

Can intelligence be reduced to a single, context-free measure, and can this measure be used to accurately predict job performance and select job candidates? To reduce a complex argument to a single sentence, maybe yes (Jensen, 1992; Ree & Earles, 1992a, 1992b; Schmidt & Hunter, 1992) and maybe no (Calfee, 1992; McClelland, 1992; Sternberg & Wagner, 1992). The very controversial use of the IQ test as the sole employment selection criterion is no mere academic exercise as companies look to the "experts" for a magic bullet that can screen upsized numbers of job seekers on downsized personnel budgets. The outcome of this controversy will have a significant impact on hiring practices in the U.S.

If intelligence could be reduced to a single objective measure, would this necessarily be a good thing? What if a set of objective conditions was established and agreed upon to determine the presence of intelligence in a machine? And further, suppose that state and corporate funding depended on a researcher's having developed an intelligent system as determined by a set of objective criteria. Would this lead to an AI systems' equivalent of the "checkbox wars" commonly fought between the marketers of shrink-wrapped software products? "Our product contains all of the necessary and sufficient conditions for general intelligent action...and more! The competition doesn't even come close." Some schools are criticized for "teaching to the test" rather than educating students. Would developers start "building to the test" in order to pass what would amount to a test of machine IQ?

Public Opinion and Public Policy

This article has presented pointers to the literature on the perception of intelligence, and where it has touched on public policy. A full review is beyond the scope of this paper. Yet despite the volume of academic research on human intelligence in the past decades, current public opinion on human intelligence and the direction of public debate has been shaped by a single book on the subject, *The Bell Curve* (Herrnstein & Murray, 1994).

It is no exaggeration to say that this book is a cultural landmark, the most widely cited and influential book on social science in decades. Attractive, well-spoken proponents for *The Bell Curve* appear with great frequency on TV and radio talk shows to argue with critics in choppy, sound-bitten debates. Its praises are sung in newspaper and magazine editorials. Well-researched academic responses to the book's many shortcomings, however, appear with little fanfare and frequently only in specialized academic journals (e.g., Sternberg, 1995). Moreover, the difference between expert opinion on the subject and public opinion is striking (House & Haug, 1995). Of course, public perceptions shape public policy, and the winners in the debate over the heritability and variability in human intelligence will decide the direction of public policy (including funding priorities) for many years. *The Bell Curve* works as well as it does by mining a vein of discontent in society that "they" are making off with "our" jobs, consuming scarce resources, and threatening our way of life. The authors present themselves as experts, present masses of impenetrable data, appeal to existing stereotypes of the "other," and do so effectively on an emotional level.

A *Bell Curve* for AI

Why should the AI community care about public reaction to a work like *The Bell Curve*? AI has seen its share of critics. There is frequently an emotional cast to the attacks on AI, but typically the most widely discussed criticisms are presented on logical grounds (Dreyfus, 1972). Rarely do the debates enter the public consciousness. Presently, however, there is a crucial element missing in the culture wars over intelligent machines. Computers may be viewed by some as an intellectual threat perhaps, but there are no widespread fears of displacement and the visceral emotional response to the "other" that underlies the current controversy over human intelligence. Evans (1979) notwithstanding, there is currently no wholesale perception that intelligent, predatory computers are breathing down human necks, poised for the kill. But that could change.

As we have seen, humans are capable of strong social responses to computers. In addition to the existing everyday frustrations over computers, there exists the potential for several spectacular computer-related disasters, including expensive, buggy IRS systems that may incorrectly process thousands of tax returns, and the impending "Year 2000" problem. Thus far, though, there is no perceived threat to livelihood, no wave of immigrant machines trumpeted in the major news media, here to take jobs and lower educational standards.

The national economy has cycles of upturns and downturns, largely indifferent to the tinkering of politicians. At the same time, a few AI researchers seek to offer convincing evidence that machines are capable of human-level, if not human-like, intelligence. If public acceptance of this view coincides with a severe economic downturn, the conditions will be present for a *Bell Curve* for AI: a wide ranging, emotional attack on the newest, politically defenseless scapegoat, the intelligent machine. The authors of AI's *Bell Curve* will hide behind a facade of statistics and events, which, while not particularly original or intellectually rigorous, will play upon existing fears of an unfamiliar "other." The public will be treated to "human interest" stories about unqualified machines which are smart (but not quite as smart as humans), self-serving, and willing to work at a fraction of the cost of good folk like ourselves. If this pessimistic scenario comes to pass, AI researchers may long for the days of the academic civility of the Dreyfuses and Searles.

If the debate over human intelligence provides any indication, the debate over AI will be fought in the media by partisans employing the devices of public relations. Expert consensus on an emotional issue such as machine intelligence may have only a limited influence on public

opinion. The public will attribute intelligence to machines based on factors beyond the control of computer scientists who are unaware of process by which humans form judgments of complex phenomena. The proponents of AI's *Bell Curve* will play to the news media's desire for conflict and a good story, and appeal to the public's negative stereotypes about computers.

Demonstrations of Machine Intelligence

The success of Deep Blue in playing chess has furnished computer scientists with an objective milestone by which to measure their progress in developing smart machines. Deep Blue has also alerted the public that something significant may be on the horizon. Intelligent entities, by one definition, show self-awareness, learn from the environment, acquire new capabilities and evolve through time (Newell, 1980). If computers begin to demonstrate these qualities to the satisfaction of all concerned, it will be hard to sell the public on the idea that computers are also devoid of self-interest, self-motivation, and self-initiated behaviors. At that point it may be too late to adopt the marketing strategy taken by one large computer manufacturing company in the 1960's that proclaimed their computers as no more than "fast, dumb servants that only do what they are told." AI researchers should consider now the possible outcomes associated with convincing the public that machines do indeed think. Of course, there may be a price of failure for being unable to satisfy public, state, and corporate demands for a substantial return-on-investment in machine intelligence. At the same time, smart machines may potentially arouse the same level of hostility in society that is usually expressed only by academics with a philosophical ax to grind. In terms of simultaneously satisfying the public's desire for a technological payoff and soothing public fears of an intellectual takeover, AI may one day find itself "between the Devil and the Deep Blue sea."

References

Amabile, T. M. 1983. Brilliant but cruel: perceptions of negative evaluators. *Journal of Experimental Social Psychology*, 19(2):146-156.

Calfee, R. 1992. Paper, pencil, potential, and performance. *Current Directions in Psychological Science*, 2:6-7.

Cook, M. 1984. The good judge of others' personality: methodological problems and their resolution. In M. Cook, (Ed.), *Issues in person perception*, 145-166. New York: Methuen.

Dreyfus, H. L. 1972. *What computers can't do: a critique of artificial reason*. New York: Harper & Row.

Evans, C. (1979). *The micro millennium*. New York: Viking.

Herrnstein R., & Murray, C. 1994. *The bell curve: intelligence and class structure in American life*. New York: The Free Press.

House, E., & Haug, C. 1995. Riding *The Bell Curve*: A review. *Educational Evaluation and Policy Analysis*, 17(2):232-272.

Jensen, A. R. 1992. Test validity: *g* versus "tacit knowledge." *Current Directions in Psychological Science*, 2:9-10.

Jones, E. E., Rock, L., Shaver, K. G., Goethals, G. R., & Ward, L. M. 1968. Pattern of performance and ability attribution: an unexpected primacy effect. *Journal of Personality and Social Psychology*, 10:317-340.

McClelland, D. C. 1992. Intelligence is not the best predictor of job performance. *Current Directions in Psychological Science*, 2:5-6.

Newell, A. 1980. Physical symbol systems. *Cognitive Science*, 4:135-183.

Petty, R. E., Cacioppo, J. T., & Goldman, R. 1981. Personal involvement as a determinant of argument based persuasion, *Journal of Personality and Social Psychology*, 41:847-855.

Ree, M. J., & Earles, J. A. 1992a. Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, 1:86-89.

Ree, M. J., & Earles, J. A. 1992b. *g* is to psychology what carbon is to chemistry: a reply to Sternberg and Wagner, McClelland, and Calfee. *Current Directions in Psychological Science*, 2:11-12.

Reeves, B., & Nass, C. 1996. *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge: Cambridge University Press.

Schmidt, F. L., & Hunter, J. E. 1992. Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science*, 2:8-9.

Searle, J. R. 1980. Minds, brains, and programs. *The Behavioral and Brain Sciences*, 3:417-424.

Sternberg, R. J. 1990. *Metaphors of mind: conceptions of the nature of intelligence*. Cambridge: Cambridge University Press.

Sternberg, R. J. 1995. For whom the bell curve tolls: A review of *The Bell Curve*. *Psychological Science*, 6:257-261.

Sternberg, R. J., Conway, B. W., Ketron, J. L., & Bernstein, M. 1981. People's conceptions of intelligence. *Journal of Personality and Social Psychology*, 41(1):37-55.

Sternberg, R. J., & Wagner, R. K. 1992. The *g*-ocentric view of intelligence and job performance is wrong. *Current Directions in Psychological Science*, 2:1-5.

Weber, B. 1996, February 19. A mean chess-playing computer tears at the meaning of thought. New York Times, A1, C8.

Weizenbaum, J. 1976. *Computer power and human reason*. San Francisco: W. H. Freeman and Co.

Zimbardo, P., & Leippe, M. R. 1991. *The psychology of attitude change and social influence*. New York: McGraw-Hill.