# Human-Like vs. Complementary Intelligence

Jacob Beal

October 26, 2004

I propose that all work on artificial intelligence can be categorized into three fields corresponding with aspects of human intelligence: animal intelligence, human-like intelligence, and complementary intelligence.

Animal intelligence encompasses everything that a rat or a chimpanzee can do, as well as less-human systems like bacterial colonies and frogs. Systems which exhibit animal intelligence struggle to survive, heal non-fatal damage, maintain homeostasis, and adapt to changing circumstances. Learning and communication is quite limited, as is long-term planning. Strong animal intelligence systems can exhibit capabilities that humans do not, like a bloodhound's ability to discriminate smells.

Complementary intelligence, on the other hand, encompasses the types of reasoning which humans find difficult like calculus, logic and chess, exhaustive learning like statistical parsing, and a wide range of other non-human techniques, from web-mining to simplex optimization. Systems which exhibit complementary intelligence engage in systematic reasoning, often requiring perfect recall from a vast memory or flawless application of rule-like transformations. Strong complementary intelligence can be used by humans for intelligence amplification, allowing individual humans to comprehend and manipulate information in ways previously intractably difficult.

What then, is the definition of human-like intelligence? There is something very important that sets us apart from our chimpanzee cousins. It seems likely to be something other than complementary intelligence, given that today's complementary intelligence systems are clearly lacking in human capabilities, despite having long surpassed us in competence at complementary intelligence tasks.

I propose that human-like intelligence centers around negotiation of the meaning of symbols. This may happen internally, as we assign labels to phenomena in our world or compare our current experiences with our memories. This may also happen externally, as we try to communicate concepts between people or as we try to establish trust or to trick one another.

In the remainder of the paper, I first speculate on how the search for human-like intelligence might have led us to focus on complementary intelligence instead. I then take an example of a human expert (a lawyer) and try to strain out the complementary intelligence that could be captured by an expert system, such that anything left may be identified as human-like intelligence. Finally, I discuss implications of this categorization for AI research, and propose directions for the field.

## 1 Why are we researching complementary intelligence?

The early days of AI were filled with heady success. Slagle's SAINT program out-scored MIT students at calculus, Wiezenbaum's ELIZA fooled people into thinking it was a person, and Newell and Simon's GPS promised to solve any logical problem. Why shouldn't we have believed that machines that were smart like people were just around the corner?

Indeed, complementary intelligence has been wildly successful. We build software which excels at playing strategic games, managing money, medical diagnosis, solving calculus problems, logical reasoning, or any number of things which humans need a intensive training to do well. Often our software can do these specialized tasks better than human experts, sometimes by orders of magnitude.

Yet the early promise bogged down, and the expert systems never generalize. Perhaps, in fact, the properties that make complementary intelligence a profitable and attractive area of study exist precisely because it has little to do with human intelligence.

I hypothesize that problems like calculus were quickly solved precisely because they are things that people are very bad at. To see why, let us consider some consequences of being bad at calculus:

First of all, the fact that humans are bad at it means that we have been forced to study the field very carefully. Conveying the necessary concepts and skills from teacher to student is difficult, and thus they must be formulated precisely, as nearly algorithmic procedures for problem solving. On the other hand, something which people are very good at doing, like communicating via natural language, is something which we can teach in an unconscious, gestural manner, yet have the student pick it up without hardly trying.

A similar effect occurs in communication between colleagues. Because we are so bad at calculus, it is not immediately obvious whether a solution to a problem is good. Thus it must be carefully formalized, with objective tests for correctness, so that colleagues who may conceptualize the problem very differently can compare notes and check one another's work.

Finally, when people are very bad at something like calculus, its complexity is of necessity limited. Moreover, it becomes regarded as something which requires a great deal of intelligence to do, because only the most intelligent people can simulate complementary intelligence well enough to perform its very straightforward tasks.

As a result, problems like calculus are ideally suited for transmission to a computer:

- All of the knowledge engineering is done: algorithms for doing calculus are taught to humans, and merely have to be adapted to be taught to computers.

- Calculus has been carefully formalized and we can easily verify whether an answer is right or wrong.
- Calculus, though intricate, is not very complex — only a small amount of knowledge is necessary to capture its full complexity.

I suspect that there also may be an unthinkability issue involved. The problems which we are easily able to brainstorm and decide need solving are those which people are bad at, not the ones in which we are so highly competent that they are transparent. Conversely, areas in which we are massively competent, such as natural language understanding, have been difficult to decompose into solvable subproblems.

We have long hoped for machines to be able to replace us at boring or difficult tasks, whether that be digging ditches or solving calculus problems, or to be more precise than us in tasks which are too dangerous to be left to mere fallible humans (mere fallible software is a different problem). These sorts of narrow problems have proven to be well-suited to complementary intelligence, not to mention extremely important in the wider world.

It's much harder to pin down what it would mean for a machine to be like us. After all, being human is not about doing any particular tasks, but rather seems to involve our ability to learn from our mistakes, adapt to the task at hand, and invent new tasks for ourselves, and indeed new classes of tasks which don't even make sense in any pre-existing paradigm.

So, as always, the key question is: what questions should research into human-like intelligence address?

## 2 What does human-like intelligence look like?

Let us attack this problem by way of examining a human expert who is hired by non-experts for her complementary intelligence skills. For example, consider a lawyer. We will strain out all of those things which are clearly complementary intelligence, and what is left will presumably be the reason why legal software has not yet replaced human lawyers.

In many ways, a human lawyer is like a well-engineered expert system. In the course of training, a lawyer is exposed to a few thousand examples of case studies and arguments based off of them, grouped into a set of thematically related clusters (e.g. tort law, contract law, etc). Law students must argue positions on the cases they study — learning by explanation which induces additional processing of the raw material, and builds expert skills in the lawyer. The actual set of information internalized by a lawyer in the end might actually be very small and may consist primarily of a few principles in every field, plus knowledge of how to search for supporting examples. A lawyer cannot in fact operate without access to complementary intelligence in the form of reference books, decisions to reread, etc.

So what is the piece of intelligence contributed by the human lawyer, which is different from the complementary intelligence provided by legal advice programs?

One key thing that the lawyer supplies is judgment of the human situation surrounding a case. Is the case not merely correct, but also something which is winnable on its merits? For example, in the field of law, it is common to have strong arguments on both sides of the case. Often, there are many ways that either side of the case can be shown to be technically correct. Leaving aside emotional affairs like jury trials, a lawyer is likely to be able to estimate how a judge will weigh these different arguments.

Lawyers are good at engaging in gaming against their opponents, and at constructing persuasive arguments, which may differ importantly from correct arguments. The persuasive quality of an argument includes not just its correctness, but how well it causes another person's beliefs to align with one's own.

A human lawyer supplies attentional and analogical skill, important because law is a huge body of regulation and precedent, and as such only elements with a high likelihood of relevance should be examined. Moreover, relevance may depend on difficult analogies, since the principles which justify a case may not be explicit until after the analogy is made.

A human lawyer also helps calm and comfort a distressed client, and should be counted on to help the client figure out where their interests lie at any point in the process. In planning strategy for a case, she must balance the client's short-term needs and long-term needs, and recommend whether to shoot for a big payoff or a small guaranteed settlement. It's important to note that this is not merely statistical judgment, even if it might be less correct than statistical judgment. If only 5 percent of a certain type of case succeeds, then statistics might lead you to believe that type of case is bad, but a human might look for explanations about what is special about that 5 percent.

There are common themes that runs through these examples. Reckoning a judge's opinions, gaming against an opponent, and interpreting a client's needs all center around estimation of another person's beliefs. Finding rules and precedent to support a case may require extending interpretations into new areas, and estimating the value of a case depends on invention of explanations for success or failure of previous cases. Constructing a persuasive argument and advising a client both center on bringing another's beliefs into alignment with one's own — though hopefully not adversarially with the client.

All of these, I propose, share one core issue: negotiating the meaning of symbols. This may be internal — inventing new ideas and shifting the definitions of pre-existing ones — or it may be between people — trying to understand another's beliefs, or to shift them to match one's own.

If this is, in fact, the case, then we should not expect symbols to have precise definitions, except in the formalized settings where AI has done so well. The more important an idea is to people, the more likely it will be subject to the push and pull of myriad people who care about its meaning. How many vital concepts are moving targets whose meanings are endlessly debated? Consider freedom, integrity, respect, trust, independence, fairness, piety, and empathy, to name just a few.

Perhaps it is even vital that our definitions are fluid and subjective. Remember that the primary competitor of a person is other people. With there are fixed definitions it is much easier to game the system than when both sides can negotiate the meaning of "fairness" in the deal which is being made.

# 3 Where do we go from here?

Assuming this hypothesis holds, it suggests a concrete research agenda towards constructing systems which exhibit human-level intelligence. I see four promising research tracks which all need to be pursued toward this goal:

- Communication. It is of primary importance to investigate how the meaning of symbols can be established and maintained, either by a person, in conversation between people, or between components of a program.
- Components. Cognitive science suggests a set of subsystems of the human mind, which could form an appropriate set of representations communicating to negotiate symbols.
- Infrastructure. The components need to be built robustly and with a flexible architecture capable of adapting to changing concepts.
- Evaluation. We need to have good standards to measure the degree to which the system exhibits human-like behavior.

As engineers, we can see that there are clear needs for all three types of intelligence. We need animal intelligence to make our systems robust and interface them with the outside world; we need complementary intelligence to solve the problems that we cannot solve alone; and we need human-like intelligence so that we can communicate meaningfully with our creations.

Note, however, that this formulation of human-like AI highlights danger as well as promise. An automated system capable of establishing trust or arguing coherently is a powerful tool capable of being perverted. Concepts like trust are moving targets in part because people are in competition, and people are constantly evolving new ways to cheat as well as new ways to verify. As we move forward, we must be careful not to create tools for which we have not developed defenses, and it is our responsibility as scientists to ensure that we are aware of the risks in this endeavor.