Subcognitive Probing: Hard Questions for the Turing Test

Robert M. French
Electrical Engineering and Computer Science Department
University of Michigan

INTRODUCTION

Alan Turing in his original article about an intelligence-game definition of intelligence seems to be making two separate claims. The first, call it the philosophical claim, is that if a machine could pass the Turing Test, it would necessarily be intelligent. I completely agree with this first claim. His second point, which I call the pragmatic claim, is that in the not-too-distant future it would, in fact, be possible to actually build such a machine. Turing clearly felt that it was important to establish both claims. He realized, in particular, that if one could rigorously show that no machine could ever pass his test, his philosophical point, while still true, would lose a great deal of its significance. He thus devoted considerable effort to establishing not only the philosophical claim but also the pragmatic claim.

Ever since his article appeared philosophers have concentrated almost exclusively on attacking or defending the philosophical claim. There are those who believe that passing the Turing Test constitutes a sufficient condition for intelligence and those who do not. As I said above, I wholeheartedly endorse the point of view that anything that could pass the Turing Test would be, without question, intelligent[2]. However, in this paper I will take issue with the pragmatic claim and argue that there is a flip side to Turing's extremely elegant test, namely, that its very capacity to probe the deepest, most essential areas of human cognition makes it, in a pragmatic sense, far too strong. The Turing Test could be passed only by things that have experienced the world as we have experienced it; the Test therefore provides a guarantee not of intelligence but of culturally-oriented *human* intelligence.

I establish this consequence of the Turing Test by proposing a class of questions, which I call *explicitly subcognitive questions*, that are intentionally designed to reveal low-level cognitive structure. Critics might object that there is something unfair about this type of question. This leads to the central idea of this paper which is that, in fact, there is *no way to distinguish questions that are subcognitive from those that are not.* To support this claim, I present another class of questions -- *implicitly* subcognitive questions -- that give every appearance of being at the cognitive level but that, in reality, are every bit as dependent on unconscious mechanisms as the initial class of explicitly subcognitive questions. In fact, close examination of some of the questions posed in Turing's original article reveals that they, too, are implicitly subcognitive. In like manner, any sufficiently broad set of questions making up a Turing Test would necessarily contain implicitly subcognitive questions. I show that it is impossible to tease apart implicitly subcognitive questions from explicitly subcognitive ones. And from this it follows that the cognitive and subcognitive levels are inextricably intertwined.

It is this essential inseparablility of the cognitive and subcognitive levels that, in a sense, undermines the Turing Test, making it too strong for its own good. As a result, it turns out to be a test for *human* intelligence, not intelligence in general. This fact, while admittedly interesting, is not particularly useful if our goal is to gain insight into intelligence in general. This seems to bring us back to the problem that Turing had hoped to sidestep by his imitation-game definition of intelligence, namely, the problem of specifying a set of necessary and sufficient conditions for intelligence. Unfortunately, there seems to be no easy way out; capturing the essence of general intelligence must be based on categorization and segmentation abilities, the ability to learn new concepts, the ability to adapt old ones to a new environment, and so on. Precisely what should be on this list and what the definition of each of its components is, as it was in 1950, still unknown.

ON NORDIC SEAGULLS

Consider the following thought experiment: Suppose that the only flying animals known to the inhabitants of a large Nordic island are seagulls. Everyone on the island acknowledges, of course, that seagulls can fly. One day the two resident philosophers on the island are overheard trying to pin down what "flying" is really all about.

Says the first philosopher, "The essence of flying is to move through the air."

"But you would hardly call this flying, would you?" replies the second, tossing a pebble from the beach out into the ocean.

"Well then, perhaps it means to remain aloft for a certain amount of time."

"But clouds and smoke and children's balloons remain aloft for a very long time. And I can certainly keep a kite in the air as long as I want on a windy day. It seems to me that there must be more to flying than merely staying aloft."

"Maybe it involves having wings and feathers."

"Penguins have both, and we all know how well they fly . . ."

And so on. Finally, they decide to settle the question by, in effect, avoiding it. They do this by first agreeing that the only example of objects that they are absolutely certain can fly are the seagulls that populate their island. They do, however, agree that flight has something to do with being airborne and that physical features such as feathers, beaks, and hollow bones probably are superficial aspects of flight. On the basis of these assumptions and their knowledge of Alan Turing's famous article about a test for intelligence, they hit upon the Seagull Test for flight. The Seagull Test is meant to be a very rigorous sufficient condition for flight. Henceforth, if someone says, "I have invented a machine that can fly," instead of attempting to apply any set of flight-defining criteria to the inventor's machine, they will put it to the Seagull Test. The *only* things that they will certify with absolute confidence as being able to fly are those that can pass the Seagull Test. On the other hand, they agree that if something fails the Test, they will not pass judgment; maybe it can fly, maybe it can't.

The Seagull Test works much like the Turing Test: Our philosophers have two three-dimensional radar screens, one of which tracks a real seagull; the other will track the putative flying machine. They may run any imaginable experiment on the two objects in an attempt to determine which is the seagull and which is the machine, but they may watch them only on their radar screens. The machine will be said to have passed the Seagull Test for flight if both philosophers are indefinitely unable to distinguish the seagull from the machine.

An objection might be raised that some of their tests might have nothing to do with flight. They would reply: "So what? We are looking for a sufficient condition for flight, not a *minimal* sufficient condition. Furthermore, we understand that ours is a very hard test to pass, but rest assured, inventors of flying machines, failing the Test proves nothing. We will not claim that your machine *cannot* fly if it fails the Seagull Test; it may very well. However, we, as philosophers, want to be absolutely certain we have a true case of flight, and the only way we can be sure of this is if your machine passes the Seagull Test."

Now, of course, the Seagull Test will rightly take bullets, soap bubbles, and snowballs out of the running. This is certainly as it should be. But helicopters and jet airplanes -- which *do* fly -- would also never pass it. Nor, for that matter, would bats or beetles, albatrosses or hummingbirds. In fact, under close scrutiny, probably only seagulls would pass the Seagull Test, and maybe only seagulls from the philosophers' Nordic island, at that. What we have is thus not a test for flight at all, but rather a test for flight as practiced by a Nordic seagull.

For the Turing Test, the implications of this metaphor are clear: the Turing Test admits of no degrees in its determination of intelligence, in spite of the fact that the intuitive human notion of intelligence clearly does. Spiders, for example, have little intelligence, sparrows have more but not as much as dogs, monkeys have still more but not as much as eight-year-old humans, who in turn have less than adults. If we agree that the underlying neural mechanisms (e.g., Hebbian learning) are essentially the same across species, then we ought to treat intelligence as a continuum and not just as "something that only humans have". It is especially important in the study of artificial intelligence that researchers not treat intelligence as an all-or-nothing phenomenon.

SUBCOGNITIVE QUESTIONS

Before beginning the discussion of subcognitive questions, I wish to make a few assumptions that I feel certain Turing would have accepted. First, I will allow the interrogator to have an assistant. I also want to make explicit an assumption that, in Turing's article, is tacit, namely that the human candidate and the interrogator (and, in this case, her assistant) are from the same culture and that the computer will be attempting to pass as an individual from that culture. Thus, if ever the computer replies, "I don't speak English" or something of the sort, the interrogator will immediately deduce, rightly, that the other candidate is the human being. Finally, while I believe that it is *theoretically* possible to build a machine capable of experiencing the world in a manner indistinguishable from a human being, I will assume that no computer is now, or will in the foreseeable future be, in a position to do so.

I will designate as *subcognitive* any question capable of providing a window on low-level (i.e., unconscious) cognitive structure. By "low-level cognitive structure", I am referring to the subconscious associative network in human minds that consists of highly overlapping activatable representations of experience. This is the level currently being explored by new approaches to cognitive modelling.[3]

The first type of subcognitive questions I will consider consists of those that I call *explicitly* subcognitive. I have chosen this name because these questions are explicitly designed to reveal low-level cognitive structure (and I think everyone would agree that they undeniably do so). I will respond to the anticipated objection that these explicitly subcognitive questions are unfair by following up with another set of questions that seem, at first glance, to be at a higher cognitive level than the first set. These questions will turn out, under closer examination, to be subcognitive as well. I will conclude with a final set of questions that seem for all the world to be innocent high-level cognitive questions but that will be just as hard as the others were for the computer to answer in the way a human would.

ASSOCIATIVE PRIMING

This first set of questions is based on current research on associative priming, often called semantic facilitation. The idea is the following: Humans, over the course of their lives, develop certain associations of varying strength among concepts. By means of the so-called lexical decision task it has been established[4] that it requires less time to decide that a given item is a word when that item is preceded by an associated word. If, for example, the item "butter" is preceded by the word "bread", it would take significantly less time to recognize that "butter" was a word than had an unassociated word like "dog" or a nonsense word preceded it.

The Turing Test interrogator makes use of this phenomenon as follows: She selects a set of words (and non-words), runs the lexical decision task on her assistant and records his recognition times. (This may, of course, be done prior to the start of the Turing Test. All that is required is that the interrogator come to the Test armed with the results of this initial test.) She then asks both candidates to perform the same experiment, and records their results. Once this has been done, she simply identifies as the human being the candidate whose results more closely resemble those produced by her assistant.

The machine would invariably fail this type of test because there is no *a priori* way of determining associative strengths between *all* possible concepts. It would be necessary for the machine to know (or be able to determine) all of the associative strengths between all human concepts in order to pass this test without having experienced the world as the human candidate and the interrogator's assistant had. Unless the machine had had this experience, there would be a noticeably higher degree of priming similarity between the assistant and the human candidate than between the assistant and the machine.

Now, suppose a critic claims that these explicitly subcognitive questions are unfair because -- ostensibly, at least -- they have nothing to do with intelligence; they probe, the critic says, a cognitive level well below that necessary for intelligence and therefore they should be disallowed. Suppose, then, that we obligingly disallow such questions and propose in their stead a new set of questions that seem, at first glance, to be at a higher cognitive level.

THE RATING GAME

Neologisms will form the basis of the next set of questions I will propose. Our impressions involving made-up words provide particularly impressive examples of the "unbelievable number of forces and factors that interact in our unconscious processing of even . . . . words and names only a few letters long".[5]

I will now propose the following set of questions, all of which have a completely high-level cognitive appearance:

"On a scale of 0 (completely implausible) to 10 (completely plausible), please rate:

- 'Flugblogs' as a name Kellogg's would give to a new breakfast cereal.
- 'Flugblogs' as the name of a new computer company.
- 'Flugblogs' as the name of big, air-filled bags worn on the feet and used to walk on water.
- 'Flugly' as the name a child might give its favorite teddy bear.
- 'Flugly' as the surname of a bank accountant in a W.C. Fields movie.
- 'Flugly' as the surname of a glamorous female movie star.
- etc."

The interrogator will give, say, between fifty and one hundred questions of this sort to her assistant[6], who will answer them. Then, as before, she will give the same set of questions to the two candidates and compare their results to her assistant's answers. The candidate whose results most closely resemble the assistant's will, without doubt, be the human.

Let us examine a little more closely why a computer that had not acquired our full set of cultural associations would fail this test. Consider "Flugblogs" as the name of a breakfast cereal. It is unquestionably pretty awful. The initial syllable "flug" phonetically activates (unconsciously, of course) such things as "fug", "thug", "flub", "ugly", or "ugh!", each with its own aura of semantic connotations, while the second syllable, "blog", no doubt activates "blob", "bog", and other words, which in turn activate a halo of other semantic connotations. The sum total of this spreading activation determines how we react, at a conscious level, to the word. And while there will be no precise set of associated connotations for all individuals across a culture, on the whole there is enough overlap to provoke *similar* reactions to given words and phrases. In this case, the emergent result of these activations is undeniable: "Flugblogs" would be a lousy name for a cereal.

What about "Flugly" as a name a child might give its favorite teddy bear? Now *that* certainly sounds plausible. In fact, it's kind of cute. But, on the surface at least, "Flugblogs" and "Flugly" seem to have quite a bit in common; if nothing else, both words have a common first syllable. But "Flugly", unlike "Flugblogs", almost certainly activates "snugly" and "cuddly", which would bring to mind feelings of coziness, warmth, and friendship. It certainly also activates "ugly", which might normally provoke a rather negative feeling, but, in this case, there are competing positive associations of vulnerability and endearment activated by the notion of children and things that children like. To see this, we need look no further than the tale of the Ugly Duckling. In the end, the positive associations seem to dominate the unpleasant sense of "ugly". The outcome of this subcognitive competition means that "Flugly" is perceived by us as being a cute, completely plausible name for a child's teddy bear. And yet, different patterns of activations rule out "Flugly" as a plausible name for a glamorous female movie star.

Imagine, for an instant, what it would take for a computer to pass this test. To begin with, there is no way it could look up words like "flugly" and "flugblogs"; they don't exist. To judge the appropriateness of any given word (or, in this case, nonsense words) in a particular context requires taking unconscious account of a vast number of culturally-acquired, competing associations triggered initially by phonetic resemblances. And, even though one might succeed in giving a program a certain number of these associations (for example, by asking subjects questions similar to the ones above and then programming the results into the machine), the space of neologisms is virtually infinite. The human candidate's reaction to such made-up words is an emergent result of myriad subcognitive pressures, and unless the machine had a set of associations similar to those of humans both in degree and in kind, its performance in the Rating Game would necessarily differ more from the interrogator's assistant's performance than would the human candidate's. Once again, a machine that had not experienced the world as we have would be unmasked by the Rating Game, even though the questions comprising it seemed, at least at the outset, so cognitively high-level in nature.

A VARIATION ON THE RATING GAME

If, for some reason, the critics were still unhappy with the Rating Game using made-up words, we could consider a variation on the game in which all of the questions would have the form:

"Rate **X**s as **Y**s" (0 = "could be no worse", 10 = "could be no better")

where **X** and **Y** are any two categories. Such questions give every appearance of being high-level cognitive questions: they are simple in the extreme and rely not on neologisms but on everyday words. For example, we might have, "Rate **radios** as **musical instruments**". Now, people do not usually think of radios as musical instruments, but they do indeed have some things in common with musical instruments: both make sounds; both are designed to be listened to; John Cage once wrote a piece in which radios were manipulated by performers; etc. There is therefore some overlap between the categories of **musical instruments** and **radios**. As a musical instrument, therefore, we might give a radio a rating of 3 or even 4 on a 10-point scale.

The answer to any particular rating question is necessarily based on how we view the two categories involved, each with its full panoply of associations, acquired through experience, with other categories. Other questions might be: "Rate **chocolate sundaes with nuts, whipped cream and cherries on top** as **antibiotics**", "Rate **grand pianos** as **wheelbarrows**", "Rate **purses** as **weapons**", "Rate **pens** as **weapons**", and so on. Just as before, it would be impossible to program into the machine all the various types and degrees of associations necessary to answer these questions like a human.

THE CENTRAL ISSUE

The central issue is that any reasonable set of questions in a Turing Test will necessarily contain subcognitive questions in some form or another. Ask enough of these questions and the computer will become distinguishable from the human because its associative network would necessarily be unlike ours. And thus the computer would fail the Turing Test.

Is it possible to modify the rules of the Turing Test in such a way that subcognitive questions are forbidden? I think not. The answers to subcognitive questions emerge from a lifetime of experience with the minutiae of existence, ranging from functionally adaptive world-knowledge to useless trivia. The sum total of this experience with its extraordinarily complex inter-relations is what defines human intelligence. And this is what Turing's imitation game tests for. What we would really like is a test for intelligence *in general* -- but how could we achieve this through a Turing-like test? Surely, we do not want to limit ourselves to questions like, "What is the capital of France?" or "How many sides does a triangle have?". If we admit that intelligence in general must have *something* to do with categorization, analogy-making, and so on, we will, of course, want to ask questions that test these capacities. But the only questions that probe these capacities are, as I hope to have shown, subcognitive questions -- and we have seen where *those* questions lead!

CONCLUSION

In conclusion, the imitation game proposed by Alan Turing provides a very powerful means of probing human cognition. But its very strength is also, in a sense, a weakness. Turing invented the imitation game as a novel way of looking at the question "Can machines think?". However, the Turing Test is so powerful that it is really asking, "Can machines think exactly like human beings?". And this is less interesting than the first question. The Turing Test provides a sufficient condition for human intelligence but does not address the more important issue of intelligence in general.

I believe I have shown that only a computer that had acquired adult human intelligence by experiencing the world as we have could pass the Turing Test. In addition, I feel that any attempt to "fix" the Turing Test so that it could test for intelligence in general and not just human intelligence is doomed to failure because of the completely interwoven nature of human subcognition and cognition. To gain insight into intelligence, we will be forced to consider it in the more elusive terms of the ability to categorize, to generalize, to make analogies, to learn, and so on. It is with respect to these abilities that the computer will always be unmasked if it has not experienced the world as a human being has. In the final analysis, the Turing Test, as subtle and elegant as it is, still leaves us with the need to define general intelligence in terms of these abilities.

ACKNOWLEDGMENTS

I especially wish to thank Daniel Dennett for his invaluable comments on the ideas and emphasis of this paper. Douglas Hofstadter also had a major influence on the ideas of this paper; in particular, he suggested the distinction between implicitly and explicitly subcognitive questions, as well as some of the ideas on the Rating Game and its Variation. I would also like to thank Daniel Andler, Gray Clossman, Daniel Defays, Pierre Jacob, Melanie Mitchell, David Moser, and François Récanati for their suggestions. This research has been supported by a grant from the University of Michigan, a grant from Mitchell Kapor, Ellen Poss, and the Lotus Development Corporation, a grant from Apple Computer, Inc., and grant DCR 8410409 from the National Science Foundation.

ENDNOTES

[1]Turing, Alan M. (1950), Computing machinery and intelligence, Mind, Vol. 59, No. 236, pp. 433-460.

[2] For a particularly clear defense of this view see: Dennett, D. C. (1985). "Can Machines Think?" How We Know, ed. Michael Shafto. San Francisco, CA: Harper & Row.

[3]Three different approaches that all address subcognitive issues can be found in:

Feldman, J. and F. Ballard (1982). Connectionist models and their properties. Cognitive Science, 6(3), 205-254;

Hofstadter, D. R., M. Mitchell, and R. M. French (1987). Fluid concepts and creative analogies: A theory and its computer implementation. CSMIL Technical Report No. 10, University of Michigan;

Rumelhart, D. and J. McClelland (1986) (Eds.). Parallel Distributed Processing. Cambridge, MA: Bradford/MIT Press

[4]A particularly relevant, succinct discussion of associative priming can be found in: Anderson, J. R. (1983). The Architecture of Cognition. Cambridge, MA: Harvard University Press, Chap. 3, pp. 86-125. In this chapter Anderson makes reference to the classic work on facilitation by Meyer and Schvaneveldt (Meyer, D. E. and R. W. Schvaneveldt (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. Journal of Experimental Psychology 90, 227-234).

[5] Hofstadter, D. R. (1984). On the seeming paradox of mechanizing creativity. In Metamagical Themas (pp. 526-546). New York, NY: Basic Books, Inc.

[6]Even though Turing did not impose a time constraint in his original formulation of the imitation game, he did claim that "...in fifty years' time [i.e., by the year 2000] it will be possible to programme computers . . . to make them play the imitation game so well that an average interrogator will not have more than 70 per cent. chance of making the right identification after five minutes of questioning" [p. 442]. In current discussions of the Turing Test, the duration of the questioning period is largely ignored. In my opinion, one reasonable extension of the Turing Test would include the length of the questioning period as one of its parameters. In keeping with the spirit of the original claim involving a five-minute questioning period, I have tried to keep the number of questions short although it was by no means necessary to have done so.