Müller, Vincent C. (2002), 'Communicating the same information to a human and to a machine: Is there a difference in principle?', in Konstantinos Boudouris and Takis Poulakos (eds.), Philosophy of communication: Proceedings of the 13th international conference on Greek philosophy (IAGP 13) (2; Athens: Ionia), 168-76. - http://www.sophia.de - http://orcid.org/0000-0002-4144-4957

Communicating the Same Information to a Human and to a Machine: Is There A Difference In Principle?

Vincent C. Müller

We try to show that there is no difference in principle between communicating a piece of information to a human and to a machine. The argumentation depends on the following theses: Communicating is transfer of information; information has propositional form; propositional form can be modelled as categorization; categorisation can be modelled in a machine; a suitably equipped machine can grasp propositional content designed for human communication. What I suggest is that the discussion should focus on the truth and precise meaning of these statements. However, in case these statements are true it follows that: For any act of communication that successfully transfers a piece of information to a human, that act could also transfer that piece of information to a machine.

11th February, 2002

1. Introduction

Let me confess at the outset that this is a programmatic paper. Given the limits of time and space and the size of the problem, what can be done is a study of what is *in principle* possible. Sometimes, I shall gesture towards solutions of technical problems without actually spelling out those solutions, in some cases knowing that such solutions do not currently exist. To put it positively, we shall provide an analysis of what are the issues and problems to be encountered in the possibility human-machine communication. Given that analysis I come to a particular conclusion, which, it seems to me, directly follows from the analysis given, so if there are issues these lie in the details of the explanations given below – e. g. on what is communication, what is information, what is belief in a machine, etc., further discussion must focus on the explanations. To prepare you for the outcome: I shall

argue that the answer to my title question is "no", i. e. there is no difference in principle between communicating a piece of information to a human and to a machine. "In principle" meaning that an entity or event that communicates to a human can also communicate to a machine.

The motivation for looking into this issue came from a combination of the theme of this conference "communication" with a project that I recently completed for the "Centre for the Greek Language" in Thessaloniki.[1] My task was twofold: 1) To convert a conventional printed Greek-English lexicon (Georgakas) into a tool readable and understandable (whatever that may mean) by a computer[2] and 2) to convert a Mediaeval Greek-Modern Greek lexicon (Kriaras) designed for computers, a database, into one readable and understandable (whatever that may mean) for humans. Actually, the latter *was* in principle readable for a suitably trained human, provided that s/he would put up with some interface curiosities – but I am jumping ahead. To my surprise, both projects turned out to be feasible, in fact the second proved a lot more tricky than the first (see Müller 2000).

This raises the question, how is this possible? And, of course, what exactly is it that was done here, does it really constitute the automatic conversion of a form of communication to a human to that to a machine (and inversely)?

As a first stab at an answer, it seems that there are considerable similarities between the structure of the communications of humans and machines – which is really not surprising, after all the machines were made by humans. More generally, the overall motivation for Artificial Intelligence (AI) is the hope to understand human cognition by discovering its workings and principles and teach these to a machine. In that vein, we should try to analyse how a human processes when a piece of information is communicated to him/her – in a dictionary entry or by somebody saying what date it is today. The contention of "human" or "strong" AI is that once we have understood the principles of human processing we can model them in a specific form of machine, a finite programmable symbol manipulating machine – a computer. ("Alien" or "weak" AI just tries produce the same results, but may use mechanisms other than those used by humans.) Of

---

1 Details on the web site of the Centre: www.kombos.edu.gr

2 Though I would grant that saying a database containing a lexicon "knows the meanings and etymologies etc. of words" is just as misleading as to call other high-level reference systems "expert systems" – as Winograd has pointed out long ago (Winograd/Flores 1986, 132).

course, we shall only look at a small implication of this large thesis, as far as communication is concerned.[3]

I will take a computer to be a "universal Turing machine", i. e. a machine that can compute any algorithm whatsoever, it can carry out any precisely defined sequence of steps that lead to a conclusion, any sequence of steps that can be mechanically carried out by an "idiot savant". Which specific form the machine takes is irrelevant but fur our purposes it is sufficient to think of a conventional van Neumann machine like the one on your desk, processing sequentially, storing in a central memory unit and processing in a central processing unit (in any case, neural networks can be modelled on such machines).

It must be expected that if we were in a position to model communication in a machine it would be useful for a general understanding of communication. What is less, if we understand what it would be for a machine to communicate, this would also shed light on the general theory of communication – which is the subject matter of this conference.

2. Communicating is the transfer of information in propositional form

Unfortunately, no-one at this conference provided me with a definitional approach to communication, so, in order to answer my title question, I will have to make up something – hoping that it will not defeat my purpose by inviting criticism of the sort: "What you defined as communication, of course a machine can do, but that is not *real* communication …". As a working definition, let us say that:

Communication is the transfer of information from a sender to a recipient

What is the "information" transferred here? The information has to have "that" form, i. e. it has to be information "that something is the case", e. g. "that today is the 23rd of August" or "that John would like to have dinner with Mary". In the common terminology, it says "that *p*", where *"p"* stands for a proposition, which can be expressed in various ways. Which proposition is communicated depends on a complex multitude of factors, apart from the sentence meaning (if any), including tone of voice, all sorts of contextual factors, etc. Propositions are the kinds of things that can be true or false but the information "that *p*" is *that information* irrespective of its truth or falsity – so one should not talk about information as knowledge, there can be false information. Being impressed by speech act theory

---

[3] For a more complete investigation, see Copeland 1993.

(especially Searle 1969), one might be tempted to think that one communicates not just the propositional content but also the illocutionary force, e. g. when one says "I promise to be here at five" or "You are an idiot!" I think that we should ignore this distinction here, taking both utterances as transferring information that *p* – even if this is, in the second case, the information that I am an idiot *and* that the speaker is insulting me. Doing something with a sentence is not communicating. Generally, it is tempting to think of counterexamples, but one has to keep in mind that it is no use spelling them out by saying "You missed out on the information that so-and-so, which is not propositional". If the information is "that-ish" it is propositional! (If communicating is communicating *something*, then we are looking at propositional content.)

So, assuming that we have established information to be propositional, in which case should we say that the information has been "transferred"? We cannot use Gricean ways and rely on the recognition of the intention of a sender (Grice 1957, 1967) because information may be transferred without such an intention being present, indeed against the intentions of a sender. So, we need the intention-bare notion of the recipient going into a state of belief, of "believing that *p*", because of the reception of the information. If this belief state is caused, we shall assume that the information has been transferred, i. e. communication took place if the information was intended to be transferred. Of course, this is a sufficient condition, but not a necessary one. I can communicate something to you without your getting into the state of believing it.

If we can establish that there is a condition under which one can say that "X believes that *p*" if X is a machine, then we would have established that communication took place (provided the belief was caused by the communication). So, under which condition is it true that "my computer believes that today is the 23rd"?

3. Propositional form can be modelled as categorisation

Modelling propositional content has proven one of the most difficult tasks of AI and I shall not pretend that this problem has been solved, but rather indicate that it could be solved in principle. Very generally, the state of believing that *p* can be analysed in good old Aristotelian ways as believing that the referent of subject term falls under predicate term. (To be sure, this is probably incomplete, leaving out relational logic, quantification etc.) So, the beliefs mentioned above are the beliefs that "Today's day falls under the category of days that are the 23rd of a month" and "John falls under the category of people who would like to have dinner with Mary". These categories (predicates) then fall under other categories in

some sort of hierarchy. Accordingly, I successfully communicate to you that it today is the "23rd" if you know that this number represents a *date, a day of the month*.[4]

4. A machine can be in a state of believing a proposition (a Categorisation)

A computer can "sort" incoming information in such a way, listing referents under concepts (or nodes in a network) – this day under that date, this person under that category, etc. If a computer can sort referents under concepts, then it may be in a state of "believing that *p*". A further indication in this direction would be if the machine were in a position to reproduce the belief following an appropriate request. (Of course, this is not a necessary condition for belief.) And, indeed, even my very simple PC can tell me what date it is if I "ask" in the right way. It even seems to know that this number *is a date* since it uses it in various internal workings as falling under the category "date", it knows what "later date" means, can organise events by date, dates by year, calculate the number of days between two dates, knows about the calendar system etc. You will notice that this is a version of the Turing test (Turing 1950) and has similar shortcomings in that it uses a purely behavioural criterion for the ascription of a psychological property (here: belief, in Turing: intelligence). It is for this reason that I added the remark that the machine appears to be so constructed that the relevant number internally plays the role of a date. This may be taken as an indication that the machine is not just "behaving as if" it knew what date it is.

Of course, there is a large school of thought arguing that symbols in a machine cannot be more than just syntactically manipulated, they will always lack "meaning" – for various reasons, like the absence of the right sort of causal connection to referents. As indicated above, I agree with one upshot of Searle's famous Chinese Room-Argument (Searle 1980), namely that the mere production of correct output is not sufficient for the ascription of mental states like believing (or, in Searle's case, the more ambitious notion of understanding). However, I think that given the right sorts of functional set-up, there is good reason to assume more than just syntactic functioning.

One might be tempted to say that this is really metaphorical talk, like saying "my car believes that we are driving at 140 km/h" – but it would seem that by now the burden of proof here in on the person who would want to say that there

---

[4] I shall not assume any particular view of concepts here, neither the prototype view (cf. Komatsou 1992) nor the classical view as sets with necessary and sufficient conditions for membership. See Fodor 1998 for a 'computational' critique of current theories.

is something missing here.[5] What is missing? Clearly, my simple PC is not intelligent and it has no consciousness. What may be required is the belief that the machine believes that this is a date. If this is just a simple awareness of the workings of the system then it seems difficult to see why such a "supervision system" could not be part of a future machine.

5. A suitably equipped machine can grasp propositional content designed for human communication

Now, let us assume we have made it plausible that a machine can get into a state of believing that *p* due to a suitable communication. What remains is to investigate whether it can get into this state because of an act of communication that is sufficient to bring a human into that state. Can a sender communicate information to a machine just like to a human? This is a technical problem – which does not mean that it is solvable, even in principle. Think of the many ways we use to communicate, even for such simple matters like my example what date it is. Think of the many factors relevant for understanding. You can type the date, say it, indicate it with a nod, show the dial of your watch – and as Dr. Pechriggl mentioned yesterday, you can even communicate by remaining silent, by doing nothing. If the *form* of communicating would be such that a machine could not understand it, the answer to my title question would have to be "yes", there is a difference in principle.

It is fair to say that the majority of current work in AI deals with such interface problems. A computer is stupid, in that it understands only digital information, so everything analogue, like sounds and images has to be translated into a digital signal *and* suitably analysed. A computer may have the sound file on its hard disk, but it does not know that this is a sound, this is a violin, or that it is Mozart; in a movie (which is just a mass of pixels, and then another in a new frame, and so forth) it does not even know which object is different from which and that one object is the same as a later one. So, there is a long way from having a digital camera focused on the dial of your watch to the computer seeing the number "23", and another long way from there to the understanding that this is today's date. The latter requires, amongst other things, knowledge about the

---

[5] Strangely, we seem intuitively more reluctant to ascribe belief than knowledge, even thought the latter implies the former. People more easily accept statements like "my computer believes that today is the 23rds" than statements like "my computer knows that today is the 23rd. Even less intuitive is the ascription of "understanding" even though that seems implied by both belief and knowledge as a result of communication.

world – teaching this to machines is another main task of AI. The technical problems associated with teaching computers such seemingly simple things are indeed formidable. The possibility of doing so relies on a deep understanding on what is involved in the human ability to perform these tasks – and, inversely, the aim to understand humans is frequently the motivation for attempting the teaching of a machine. Again, I can see no principled reason why a machine should not eventually be able to perform these tasks (*pace* Dreyfuss 1979), even though it might take us very long indeed to get to this point.

It may be worth mentioning that there is a possibility of responding positively to my title question, but for reasons completely different from the ones we looked at so far. Are there perhaps communications of a piece of information that work for a machine, but not for a human? Clearly, an unaided human cannot understand what is running through the network cabling of his local PC-Network or the Internet connection he may have. There is a massive interface problem for us. Theoretically, however, you can use a machine (just as the computer used a camera) to translate the electric signals into something analogue that you can at least perceive, e. g. an image on a screen. And then, if you have the appropriate training, you can receive the information that is transported there – though it might not cause any belief in you other than something of the form "computer says that the control bit of the previous packet was false". As machines become more complex and more remote from their human designers it is conceivable, I think, that the information passed between machines is only comprehensible by the use of these very same machines – you might consider this a "difference in principle" in the intended sense.

6. Summary

We have tried to show that there is no difference in principle between communicating a piece of information to a human and to a machine. The argumentation depended on the following theses:

- Communicating is transfer of information
- Information has propositional form
- Propositional form can be modelled as categorisation
- Categorisation can be modelled in a machine
- A suitably equipped machine can grasp propositional content designed for human communication

Some arguments in favour of these theses have been advanced, but there is clearly room for doubt. What I suggest is that the discussion should focus on the

truth and precise meaning of these statements. However, in case these statements are true it follows that:

For any act of communication that successfully transfers a piece of information to a human, that act could also transfer that piece of information to a machine.[6]

References

Copeland, Jack B.: (1993) *Artificial Intelligence. A Philosophical Introduction*, Oxford: Blackwell.

Dreyfus, Hupert: (1979) *What Computers Can't Do: The Limits of Artificial Intelligence.* 2nd ed., New York: Harper & Row.

Fodor, Jerry A.: (1998) *Concepts: Where Cognitive Science Went Wrong.* Oxford: Oxford University Press.

Grice, Paul: (1957) "Meaning", in: Paul Grice: *Studies in the Ways of Words*, Cambridge, Mass.: Harvard University Press 1989, 213-223.

— (1969) "Utterer's Meaning and Intentions", in: Paul Grice: *Studies in the Ways of Words,* Cambridge, Mass.: Harvard University Press 1989, 86-116.

Komatsu, Lloyd K.: (1992) »Recent Views of Conceptual Structure«, in: *Psychological Bulletin* 112, 500-526. (14-19)

Lakoff, George: (1987) *Women, Fire, and Dangerous Things, What Categories Reveal about the Mind,* Chicago: University of Chicago Press.

Müller, Vincent C.: (2000) "How do We Read a Dictionary (as Machines and as Humans)? Kinds of Information in Dictionaries Constructed and Reconstructed", in: Evangelos Dermatas (ed.): *Proceedings of COMLEX2000, Computational Lexicography,* Patras University Press, Patras, 141-144. (Also on http://www.wcl2.ee.upatras.gr/comlex2/f32.pdf)

Searle, John: (1969) *Speech Acts.* Cambridge: Cambridge University Press.

[6] I am very grateful to the participants of the conference for their very numerous stimulating questions, particularly to Gottfried Heinemann and David Hitchcock.

Searle, John: (1980) "Minds, Brains and Programs", in: Magret A. Boden (ed.): *The Philosophy of Artificial Intelligence,* Oxford: Oxford University Press 1990, 67-88.

Turing, Alan: (1950) "Computing Machinery and Intelligence", in: Magret A. Boden (ed.): *The Philosophy of Artificial Intelligence*, Oxford: Oxford University Press 1990, 40-66.

Winograd, T. A./Flores, F.: (1986) *Understanding Computers and Cognition,* Norwood, N. J.: Ablex.

VINCENT C. MÜLLER
CHAIR, DEPT. OF PHILOSOPHY & SOCIAL SCIENCES
AMERICAN COLLEGE OF THESSALONIKI
P. O. BOX 21021
55510 PYLEA
GREECE
VMUELLER@ACT.EDU