Invention, Intension and the Extension of the Computational Analogy*

Hajo Greif[1] [0000−0002−1003−7494]

ICFO, Warsaw University of Technology
mail@hajo-greif.net
https://hajo-greif.net

Abstract. This short philosophical discussion piece explores the relation between two common assumptions: first, that at least some cognitive abilities, such as inventiveness and intuition, are specifically human and, second, that there are principled limitations to what machine-based computation can accomplish in this respect. In contrast to apparent common wisdom, this relation may be one of informal association. The argument rests on the conceptual distinction between intensional and extensional equivalence in the philosophy of computing: Maintaining a principled difference between the processes involved in human cognition, including practices of computation, and machine computation will crucially depend on the requirement of intensional equivalence. However, this requirement was neither part of Turing's expressly extensionally defined analogy between human and machine computation, nor is it pertinent to the domain of computational modelling. Accordingly, the boundaries of the domains of human cognition and machine computation might be independently defined, distinct in extension and variable in relation.

Keywords: Computer models · Artificial Intelligence · Intensional vs extensional equivalence · Turing computability · Limits of computation.

1 The problem: delimiting cognition and computation

Much of the presumed contrast between human and machine abilities with respect to computation rests on the assumption that the constraints on machine abilities lie in the formal nature of the operations of digital computers, and that the cognitive privilege of human beings lies in those aspects of cognition which are not covered by the computational properties of those machines and thereby escape their specific formal constraints. It will be impossible, the argument goes, to provide computer-based models of these aspects of cognition either because they are fundamentally not computational in nature or because they are *de facto* not amenable to models based on Turing-computable functions.[1]

* This is a slightly revised and extended version of a Computability in Europe conference submission. The author extends his gratitude to Paula Quinon, Paweł Stacewicz and his other colleagues at ICFO.

[1] As a disclaimer, I have to add that I am an outsider to the debates in the philosophy of mathematics in general and computation in particular. My main research area

Both the computer analogy and its limitations are grounded in Alan Turing's (1936) endeavour of modelling his theoretical machine, the Logical Computing Machine (LCM, later known as the Turing Machine), on the behaviour of human computers: breaking down complex mathematical operations into elementary arithmetical routines that can be accomplished with only a modicum of mathematical skills. Hence, only a small subset of human cognitive abilities is involved in computing, thus conceived. On this level, and notwithstanding the possibility that all higher-order abilities are ultimately realised by lower-order processes, human computation is distinguished from higher-order human mathematical skills precisely by not involving inventiveness, intuition, creativity or any other hard-to-formalise abilities. Moreover, not the inner nature and hence the mode of realisation of computational skills in human beings are addressed by the model but only the analogy in rule-governed behaviour between human computers and the LCM. Given the local and restricted analogy to human cognitive abilities, the question is whether or not concrete machines based on the LCM, i.e., digital computers, might in principle become able to accomplish mathematical and other cognitive tasks that lie beyond the domain of that computational analogy, so that they may cover all domains of (human) cognition.

For the sake of argument, I will work with the assumption that the domain of human cognition is larger than what is covered by that restricted analogy, whereas, by the same token, the domain of machine computation remains confined to that restricted analogy. The burden of proving otherwise lies on the project of Artificial Intelligence. It amounts to demonstrating that the domain of human cognition can in fact be captured by that prima facie restricted analogy, either because all aspects of human cognition can be shown to be Turing-computable or because the computational method can be made to transcend the confines of Turing-computability. Either way, the analogy is taken to be one of equivalence between the operations involved in human and machine computation in the first place, independent of whether the analogy is positive or negative.

More implicit in these debates is the question of the nature of the requisite equivalence: if it is supposed to hold between the internal operations of the systems under comparison, it will be intensional in kind, and hence will require that their concrete internal structure and operations are identical in every relevant respect. Intensionality is the quality of a couple of terms of not merely referring to the same subject matter but of doing so in the same way, on the grounds of the same internal states of the persons or systems uttering them. (This basic distinction in philosophical semantics goes back to the sense / reference dichotomy introduced by Frege 1892, and is has been central to the notion of an autonomous domain of intentional phenomena ever since.) However, if intensional equivalence is required in the present context, it will likely never come to pass, since the – neuronal, analog versus electronic, digital – modes of realisation are and will remain different in kind even if and when referring to the same subject matter. Conversely, if the equivalence merely has to hold between the formal descriptions

---

is the history and philosophy of Artificial Intelligence (AI) and cognitive science. Hence, I will have very little to say about the mathematical matters involved here.

of the outward effects of the operations involved, and hence on the level of their functions, it will be extensional in kind. It can therefore be demonstrated to hold with respect to all those levels of human cognition which can be subsumed under the same kind of formal description. (For the notion of intensionality in mathematics and computation, see Feferman 1985 for foundational work and Antonutti and Quinon in press for contemporary debates.)

On this background, the argument against equivalence is either that it cannot be accomplished at all because it cannot be intensional, or that it can be extensional but then will remain restricted to those domains of human cognition which can be subject to formal descriptions – which presumes that there are other domains to which this condition does not apply. I will address the second argument first before providing more detail on the intensional/extensional distinction, as it will have a bearing on the latter.

2 The domain of human invention

Besides the well-rehearsed arguments pro and contra cognition being an intrinsically and irreducibly embodied phenomenon, a paradigm of the debates concerning the potential non-formal constituents of human cognition is the question of mathematical intuition and invention (which I, for the sake of argument, will treat as one package). Against a view of cognition which equates cognitive with computational processes in their entirety, one can provide at least two distinct cases for the relevance of invention and intuition to mathematics, with various possible shades between them:

i.1 According to the weak, 'hypothetico-deductive' version (which might well be compatible with scientific and mathematical orthodoxy), mathematical invention and intuition are analogous to the role of invention and intuition in the empirical sciences as conceived of by Einstein, Popper and other non-inductivists, and as such are well-circumscribed in their roles. They are important to the context of discovery, where hypotheses might be formed rather freely and informally, and even independent of established standards of rationality (Kekulé famously claimed he dreamt up the benzene ring), and then put to the test. The testing belongs to the context of justification though, where fundamental principles apply that are not subject to human inventiveness and that operate in more determinate fashion. Hence, there are constraints on human invention, either in the empirical world or in mathematical principles.

i.2 According to the strong, 'constructivist' version, there are no principled limits to mathematical invention and intuition, as virtually all mathematical principles are human inventions, apart from the law of non-contradiction and a few other elementary logical principles at most. Possibly even as far down as number concepts (which, however, can be found to some extent in some animals, see Dehaene 2011; Fabry 2018), but certainly on the level of higher-order principles, mathematics is a human invention. (For example,

it has been a matter of debate whether the number zero was discovered or invented by Indian mathematicians.) This view has an analogy in constructivist approaches to the empirical sciences, according to which everything that epistemologically matters, including the conceptions of the objects under investigation and their properties, are subject to human invention. Either way, there are no principled constraints on what human ingenuity could accomplish, only practical ones.

Hence, if there are principled constraints on human inventiveness according to i.1, these can be described or circumscribed by a set of fundamental principles that can be expressed in formal terms. This does not imply that human inventiveness is ultimately equally reducible to formal descriptions but that there are boundary conditions to it that can be thus described. It thereby becomes bound to a certain domain whose extension can be determined in principle. There may well be non-formal constituents of human cognition, and these may well resist computational modelling, but their kind and scope will be circumscribed by principles that may in turn be at least partly amenable to computational modelling. If, however, there are no principled constraints on human inventiveness according to i.2, there will neither be a need for circumscribing the boundaries of its domain in formal terms, nor will there be a possibility to do so. Any and all formal terms and principles that human thinkers could devise will then arise from intrinsically non-formal origins.

With respect to computational models of human cognition, these two approaches will have quite distinct prima facie implications. The models might be powerless with respect to providing insight into human cognition in i.2 precisely because human inventiveness is so powerful. Provided enough time, resources and the right sort of invention, they could be made to fit in any way that comes to pass. The questions would then be whether they still were to be *computational* models, given the open-ended nature of human inventiveness, and whether they still were to be *models*, given that the specific epistemic quality of models in science lies in creating partial, constrained analogies between distinct systems that help to generate predictions and theories (this is the classic view of models in science established by Hesse, 1966).

In i.1, any model of human cognition will be subject to the set of principled formal constraints identified for the domain of human invention at a minimum, and at at a maximum will remain constrained to Turing's computational analogy. Either way, the type of equivalence relation involved will be open to debate. Extensional equivalence might be admitted on the level of the restricted analogy between machine computation and the sub-domain of human cognition that passes as computation on Turing's terms, but it will only do so on the functional, not the realisation level. It might also be admitted to the modelling of the principles that delimit the domain of non-formal constituents of human cognition. However, intensional equivalence would remain out of reach unless cognitive processes could be proven to *be* computational processes, which would undermine the very premiss of i.1 though and amount to strong computationalism.

3 The limits of computation

The implications of i.1 and i.2 on questions of computational modelling are merely *prima facie* because they follow by implicature and association rather than logically. Neither of the above accounts says anything about the computational analogy involved. Instead, both presuppose that it is tightly limited by default. They then offer contrasting resolutions to those limitations. The reasoning is that, if there are constituents of human cognition which are not formal in nature, a computational model of these constituents, for being formal in nature, will be inadequate. However, first, modelling relations are not identity relations but partial analogies to begin with. Second, the partialness of these analogies may change over time, precisely because new modelling methods, formal or other, might be invented without being (fully) predictable. Hence, i.1 and i.2 might be right about human cognition without being right about computation.

To make this point clear, I suggest to resort to a reconstruction of Turing's original definition of computation (which he famously never made fully explicit):

c.1 The domain of computable functions is exhausted by the functions that are 'effectively calculable' in such a way that they can be solved, in principle, by a logical computing machine (LCM), as described in Turing (1936).
c.2 A LCM comprises of a finite set of symbols, a finite set of possible states, a transition function and a potentially infinite memory.

Hence, everything that an LCM or 'Turing Machine' can solve is computable. This is has often been taken to amount to the claim that everything that *any* machine can solve is Turing-computable, and hence that every machine is computationally equivalent to a Turing Machine. Most notably, Robin Gandy's "Thesis M" (1980), which is derived from Turing's thesis, states that "whatever can be calculated by a machine can be calculated by a Turing machine" (Copeland, 2009, p. 10) and, conversely, "anything that a machine can do is computable" (Hodges, 2008, pp. 86-7). Either way, this is a notably stronger claim than Turing-computability as originally conceived, which did not concern the principled abilities or limitations of machines *qua* machines.

Still, Thesis M has given rise to a "maximal" programme of reasoning about the nature of computation, which has been continued towards the notion of conceivable machines that could, in principle, solve *more* than the set of functions computable by LCMs. This notion raises question of whether or not we should still refer to whatever goes on in those more powerful machines as computation, and why we should or should not do so. If we take Turing's definition at face value, computability is always relative to the design of his LCM, and it would take some justification to argue, with reference to Turing, for something as computation that is not covered by his definition.

However, to complement Turing's own restrictive concept of computability, he introduced notion of an 'oracle', which could solve all the functions that an LCM *cannot* solve. Still, there are two divergent interpretations of what an oracle could be and accomplish (as Turing was notoriously vague on this point, too):

o.1 If a machine is necessarily restricted to computable operations in terms of Turing's LCM, there will be principled limitations on conceiving and building a machine that could solve any non-computable problem. If there is an oracle, it will not be a machine (Hodges' interpretation).
o.2 If a machine were possible that solves non-Turing-computable functions, Turing's oracle could become a real machine in principle. There will be no essential limitations on what a machine could possibly accomplish, but that machine would not be a LCM (Copeland's interpretation).

Matching this distinction against the distinction between different readings of the role of invention and intuition in mathematics, we end up with two contrasting 'no principled limitations' claims concerning human inventiveness (i.2) and machine abilities (o.2) respectively, and with two less contrasting 'principled limitations' claims (i.1 and o.1). The parallelism of these claims is superficial though, since the limitations or the lack thereof may play out differently on the human and machine sides, for want of a necessary connection between them. If we juxtapose the positions discussed under "i.n" and "o.n", the following landscape of hypotheses emerges, with not all of its elements being equally plausible:

h.1 If mathematical principles are human inventions (as in i.2), and if human beings could build a machine that solves non-Turing-computable functions that captured non-Turing-computable elements of the human mind (as in o.2), this machine would not be a Turing Machine but an oracle-machine. Strong AI would be possible but it would not be Turing-Machine-based AI.
h.2 If mathematical principles are human inventions (as in i.2), and if human beings could not possibly build machines that solve non-Turing-computable functions that captured non-Turing-computable elements of the human mind (as in o.1), it might still be possible in principle to invent other, yet unspecified but non-mechanical routes to solving those functions. There would be no machine-based route to Strong AI.
h.3 If at least some fundamental mathematical principles are not human inventions (as in i.1), and if the principled constraints on what a machine could provide in terms of computational solutions are restricted to Turing-computability, the constraints in question will be strict for machines (as in o.1) while being differently and less narrowly defined for humans. Strong AI would not be possible. We can only build Turing Machines.
h.4 If at least some fundamental mathematical principles are not human inventions (as in i.1), and if the principled constraints on what humans and machines alike could provide in terms of computational solutions were wider than the constraints of Turing-computability (as in o.2), the constraints on machine models might coincide with the boundaries imposed on human inventiveness. Strong AI would be possible, but an oracle-machine could not be more powerful than what human beings can accomplish.

I have again left out strong computationalism, according to which, on the most extreme interpretation, human beings could invent a machine (either of the Turing or of the oracle kind) that would potentially be more powerful than whatever

human cognition could achieve. Second in the order of implausibility is h.2, as it envisions the possibility of inventing artefacts that are not machines in any known meaning of the word and that are capable of problem-solving in equally unforseeable ways. But then, Turing (1950, p. 442) made a point of using his LCM for pushing the envelope of the meaning of the word "machine" at a time when there were no computing machines.

4 Concluding remarks

I have no proof or other formal conclusion to end on but merely one observation, a morale, another observation and yet another morale: First, the relation between the limits of computation and the limits of human inventiveness remains an open question, with each side of the equation having to be solved independently.

Second, it will be worthwhile to expressly acknowledge and address the relation between human and machine abilities as an open question, and as multifaceted rather than as a strict dichotomy. Any possible decision for one position or another will have rich and normatively relevant implications. On most of the more tenable accounts outlined above, the domains of human cognition and machine computation will be distinct in kind and extension, but this will be not a matter of a priori metaphysical considerations but of empirical investigation and actual, concrete human inventions.

Third, whatever the accomplishments of AI are and may come to be, intensional equivalence is not going to come to pass. In fact, several of the classical philosophical critiques of AI build on the requirement that the same cognitive functions would have to be accomplished in the same way in machines as in human beings for AI to be vindicated. Even if questions of AI are not involved, different kinds of computing machines – for example analog, digital and quantum computers – might provide identical solutions to the same functions, but the will do so in variant ways. Hence, intensional equivalence will remain out of reach here, too.

Fourth, intensionality is an interesting and relevant concept in mathematics and partly also in computing, to the extent that one is concerned with the question of what mathematical objects are to human beings (which was the explicit guiding question for Feferman 1985). However, intensional equivalence might prove to be too much of a requirement when it comes to comparing realisations of computational processes in human beings and various types of machines. Extensional equivalence will have to suffice. It might become a more nuanced concept once we define the analogies involved with sufficient precision and move beyond the confines of pure Turing-computability. After all, Turing's computer analogy builds on extensional equivalence between human and machine operations. This kind of equivalence and its possible limitations are essential to the very idea of computer modelling. This leaves open the possibility of other relations of extensional equivalence to hold between different types or levels of systems, computational or other.

References

Antonutti, M. and P. Quinon, eds. (in press). *Intensionality in Mathematics. Synthese* Special Issue. Berlin/Heidelberg: Springer.

Copeland, B. J. (2009). "The Church-Turing Thesis". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Spring 2009. Stanford: The Metaphysics Research Lab, html. URL: `http://plato.stanford.edu/archives/spring2009/entries/church-turing/`.

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics*. 2nd ed. Oxford/New York: Oxford University Press.

Fabry, R. E. (2018). "Turing Redux: Enculturation and Computation". In: *Cognitive Systems Research* 52, pp. 397–808.

Feferman, S. (1985). "Intensionality in Mathematics". In: *Journal of Philosophical Logic* 14, pp. 41–55.

Frege, G. (1892). "Über Sinn und Bedeutung". In: *Zeitschrift für Philosophie und Philosophische Kritik* 100.1, pp. 25–50.

Gandy, R. (1980). "Church's Thesis and Principles for Mechanisms". In: *The Kleene Symposium*. Ed. by H. J. K. Jon Barwise and K. Kunen. Vol. 101. Studies in Logic and the Foundations of Mathematics. Elsevier, pp. 123–148. DOI: `10.1016/S0049-237X(08)71257-6`.

Hesse, M. B. (1966). *Models and Analogies in Science*. Notre Dame: University of Notre Dame Press.

Hodges, A. (2008). "What Did Alan Turing Mean by "Machine"?" In: *The Mechanical Mind in History*. Ed. by P. Husbands, O. Holland, and M. Wheeler. Cambridge/London: MIT Press, pp. 75–90.

Turing, A. M. (1936). "On Computable Numbers, with an Application to the Entscheidungsproblem". In: *Proceedings of the London Mathematical Society* s2-42, pp. 230–265.

— (1950). "Computing Machinery and Intelligence". In: *Mind* 59, pp. 433–460.