SOME THOUGHTS ON AUTOMATION AND MATHEMATICAL RESEARCH

AKSHAY VENKATESH

> *The deeper one digs the spade, the harder the digging gets; maybe it has become too hard for us unless we are given some outside help, be it even by such devilish devices as high-speed computing machines.* (Weyl, [6].)

Introduction. In 2017 Deepmind's *Alphazero* taught itself chess and Go "overnight," surpassing human performance and apparently reconstructing a good part of accumulated knowledge about chess openings. We will consider a thought experiment:

> What if, in ten years, "Alephzero" (written $\aleph(0)$) does the same for mathematics?

"Mathematics" for the purpose of this essay means "research in pure mathematics." Our starting point is to imagine that $\aleph(0)$ teaches itself high school and college mathematics and works its way through all of the exercises in the Springer-Verlag *Graduate Texts in Mathematics* series. The next morning, it is let loose upon the world – mathematicians download its children and run them with our own computing resources. What happens next – in the subsequent decade, say?

This is indeed a thought experiment, for it is clearly unrealistic: By restricting our horizon to ten or twenty years in the future, we allow ourselves to consider the question in isolation from the social changes that would likely accompany this kind of technological advance, and also allow ourselves to avoid thinking about more extreme types of machine intelligence – we model $\aleph(0)$ as a power tool and not as a sentient collaborator. Nonetheless I have found the exercise to be clarifying.

We may comfort ourselves with the thought that, in reality, the premise is so far in the future that we need not think about it. But if we allow even a remote possibility that this might happen in twenty years – the timescale between commencing an undergraduate degree and obtaining tenure – it certainly merits us grappling with the possibilities. I suggest that:

- Human mathematics may go on as before in many respects, just as many other professions have adapted to automation. Indeed, the resulting mathematics will be inestimably more powerful than ours, in the sense that its ability to solve any specific question will be vastly greater.
- However, the resulting field will be greatly altered; its central questions and values will be very different from those to which we are accustomed, rendering it all but unrecognizable to us.

This is a writeup and expansion of a talk I gave at the IAS in November 2021 as part of an ongoing interdisciplinary seminar examining some of the impacts of machine learning.

The main point I want to make here is that the mechanization of our cognitive processes will enhance our ability to do mathematics but also will alter our understanding of what mathematics is. We cannot meaningfully assess the first point without taking into account the second. To look at it seriously we must examine, at a minimum, the effect of automation on those processes by which our field decides which questions are interesting and fruitful; as practitioners we rarely stop to think about these, but, even setting aside our current purpose, there are many reasons not to leave the examination of such matters entirely to historians and sociologists of science.

In the remainder of the essay, I will discuss how value and consensus is constructed and maintained in current research mathematics, and then consider how $\aleph(0)$ will affect some of these processes.

Preliminary observations. We should begin by observing that human mathematical research is in no danger of being killed.There is a very large gap between the ease of asking a question and the difficulty of answering it; and for a meaningful notion of human research it is sufficient that we understand the questions but cannot solve them readily.

It is tempting to wonder about the specifics of $\aleph(0)$'s capabilities. Will it be able to visualize higher dimensions? Will it produce proofs that are displeasing, or even oracular insight without proofs? Will it surpass us at *all* mathematical reasoning tasks (a scenario that we should certainly not dismiss)? Indeed, it is very hard to imagine the exact structure of post-$\aleph(0)$ mathematics without some understanding of such issues. But we can still hope to obtain insight without such details, simply by thinking of extreme versions of commonplace phenomena. For example, many consequences of the development of $\aleph(0)$ will resemble the consequences of a very large increase in the number of working mathematicians. The experience of $\aleph(0)$ producing alien insight without proof would also not be wholly foreign to us, for our colleagues in physics departments have done this for a long time, and with less electricity consumed.

It is similarly irrelevant to our current purpose to know whether $\aleph(0)$ can enter mathematical realms that are essentially beyond our comprehension. We will regard this as the proverbial tree falling in an unpopulated forest, i.e. we are interested only in the effect on humans.

Value and consensus in mathematics. There are infinitely many mathematical problems, and a finite number of mathematicians. Very few mathematicians substantively interact with a typical problem, and conversely a single mathematician can be aware of only a small part of the mathematical landscape. By what mechanism, then, does it happen that there is a substantial measure of consensus on what the important problems are, at least at a given time, and even stronger consensus on who is doing important work? I don't mean to suggest, of course, that we mathematicians have anything near unanimity on such issues. However, my impression is that we have much more of it than other academic fields.

The valuation mechanism is fundamentally important, because it constrains with an iron, if invisible, hand the mathematics we can feasibly do. It is responsible for selecting what we are exposed to in talks, seminars and papers, and for incentivizing some questions over others. In a sense, it defines what mathematics is at any given time. So it is crucial to examine carefully how this value structure evolves. The points I am about to make are very simple ones, instinctively grasped by mathematicians in our working lives, but they are not often enunciated explicitly.

There are some obvious mechanisms that influence the construction of value:

(a) External validation (for example, the influence of applied fields such as cryptography or fluid mechanics);
(b) Processes that direct our attention (e.g. seminars, conferences, journals, prizes, influence of individual charisma, social media);
(c) Infrastructure (e.g. the organization of the educational system, the hiring process, and the grant process);
(d) Aesthetic considerations.

We shall assume these mechanisms will evolve slowly in relationship to the transition we want to study, and so we will not discuss them. This is clearly not entirely realistic and point (b) is particularly important, both because it has evolved very rapidly in recent times (e.g. through the creation of giant online seminars), and because it mediates the processes discussed below.

In any case, (a) — (d) miss a crucial part of the picture, because they are not specific to mathematics, and I think they do not adequately explain why mathematics should have a *higher* level of consensus than other academic fields. There is one feature of mathematics that stands out: it has distinguished a specific class of scholarly communication (proofs) which are *defined* by the fact that they should induce uniform agreement about their validity, without any need for replication.[1] It is reasonable to suppose that our elevated level of broad consensus is eventually derived from our much higher level of consensus on the narrow issue of validity of proof. I will assume this is so, although it is by no means obvious; to investigate this point further, it would be useful to compare with fields such as physics, economics and computer science where proof plays a substantial but less central role. In any case it becomes important to study how consensus might propagate from a restricted setting to a broader one.

There are many situations, such as the price mechanism in a free market or the Elo rating system of chess, where information is propagated through a network through repeated local transactions, thereby arriving at a consensus even when individual actors have only local information. I suggest that a similar mechanism, which we could informally call

[1]In fact, *in practice*, the correctness of mathematical proofs is at least partly maintained by a process of replication, and it is a interesting topic of current discussion how close modern proofs are to being formally valid. However, all that is important for us here is that a proof is generally understood to mean an argument compelling consensus.

(e) free trade in ideas,[2]

is a crucial component of the valuation mechanism in mathematics. I will describe it as a Bayesian process of updating our mental landscape of mathematics and mathematicians as we receive information about it. Models of this type have been studied extensively in different contexts, see, for example, [3, 5] for examples from computer science and cognitive science respectively.

Tautologically, the value we assign to a work of mathematics is purely subjective, in the sense that it depends solely on the perception of that work, and not on any objective quality. Through what means is a work of mathematics perceived by other mathematicians? The size and complexity of modern mathematics means that most papers are almost incomprehensible to us; our opinion of them can then only repeat that of others. The only people who can be involved in the formation of opinion about a given paper or a given question are those who interact with it in some way. Now, the set of people who study the details of any argument themselves is very small; a much larger group acquire, instead, an awareness of its relationship to other existing work. This can be acquired quite incidentally, e.g. through attending talks, reading or refereeing papers, reading or writing recommendation letters, and other less formal methods. Let us, proceeding by way of example, examine how such awareness of the relationship between different works can shape opinion.

Suppose that we learn of a relationship between two conjectures in our field:

$$\text{conjecture } X \implies \text{conjecture } Y. \tag{1}$$

This could mean that (i) conjecture X is more important than we thought, or that (ii) conjecture Y is easier than we thought. In practice we decide (to some extent unconsciously) according to the prior uncertainty of our beliefs: if Y is a conjecture of long standing, option (i) is more likely, and if X is a conjecture of long standing, option (ii) is more likely. Nor does X need to imply Y for this conclusion - they need only be linked in some substantive way. A similar situation occurs if

$$\text{mathematician } A \text{ proves conjecture } Y; \tag{2}$$

this is possible evidence that either A is a good mathematician, or that Y is an easy conjecture, and in practice we again choose in a fashion dependent on our prior information. In either of the situations (1) or (2), our views and uncertainty about *both* interacting parties are altered.

The intellectual activity in a field involves innumerable interactions of this general type. (It is a gross oversimplification to reduce mathematics to a collection of events of type (1) and (2), but we will adopt this very crude model for our current discussion, keeping in mind its obvious limitations.) The endless iteration of the resulting value negotiations is an important means by which the value of problem

[2]This phrase, suggesting a market metaphor for an intellectual process, appears in the dissent of the justice Oliver Wendell Holmes in a famous decision of the United States Supreme Court [1]; that text continues "the best test of truth is the power of the thought to get itself accepted in the competition of the market."

X is established within the "vicinity" of X, i.e. among those people to whom problem X is visible, and is also perhaps the dominant means of establishing a status hierarchy among workers in that community. The specifics of how this mechanism work are of course heavily influenced by what defines "visible," in particular the processes mentioned in (b). Now, although two observers A, B in the same field do not observe the same interactions and they do not in general interpret identically those that they do both see, there is nonetheless a substantial fraction on which they agree, precisely because of the concept of rigorous proof. This reduces the discrepancy between the value systems deduced by A and B.

To spell out: when will a new conjecture X acquire a high value in this model? This will be so, to the greatest extent, if both of the processes (1) and (2) just described raise its status, which is to say:

(a) It is *difficult*: many people try to solve X and fail.
(b) It is *central*: X is linked with many other questions of (prior) importance.

An interesting empirical study of the relative status of different research fields within mathematics has been carried out by Schlenker [4]. He examines which subfields of mathematics have the most "prestige," this notion being defined via bibliometrics, prizes and departmental rankings; to explain his results, he hypothesizes that fields of high "prestige" are distinguished by a *focus* on a *small number* of *central* questions.

How does this hypothesis relate to our discussion? We just noted that our simple model predicts the role of *centrality* in determining status. The function of *small number* is that problems require many repeated attempts at solution (strictly, many repeated *visible* attempts) to certify their difficulty. This is only possible when the number of workers is large relative to the number of questions.

But then – why do some fields have fewer central questions than others? I cannot see any meaningful or intrinsic sense that one field has "fewer" problems than another. Partly the emergence of central questions may reflect the structure of the mathematics itself, which is very difficult to quantify, but a readily visible factor is the extent of barriers to working on new problems. Where such barriers are low (as, for example, in combinatorics)[3] the set of problems under investigation can be relatively large in comparison to the number of workers in the field.

It is also interesting to consider failures of consensus, which may arise because different observers see different parts of the network. Consider, for example, problems X that are common to two fields C, D which otherwise have little overlap. Observers from field C and those from field D then see X within entirely different "contexts" and its importance may be perceived rather differently within the two fields. This can even happen when field D is an offshoot of field C, or potentially when D and C are the same field at different times. Increases in the overlap of C and D would probably lead to equalization.

[3]A colleague of mine, in reading this, felt that it might be interpreted as demeaning combinatorics. My intention is in fact quite the opposite. If anything I hope that analyzing the origins of our conceptions about "depth" will make us think more critically about those conceptions.

I have attempted here to mechanistically model some part of how valuation in mathematics operates in practice, but I am not advocating any position on how it *should* work. To discuss this, we would first need to clarify what the goals, internal and external, of mathematics research are; such a discussion can obviously go on without end – which is fortunate, because in our post-$\aleph(0)$ existence the fundamental role for humans may be exactly to carry on this conversation.

The impact of mechanization. We have offered a rough model of part of our valuative mechanism via (Bayesian) interaction in a network of mathematicians and problems.We now consider how $\aleph(0)$ will affect this network and alter the resulting outcome.

Perceived difficulty is, as we have seen, an essential component of our construction of value. No matter the specifics, $\aleph(0)$ will alter our ability to solve questions and therefore our perception of their difficulty. The parts of the mathematical process that can be speeded up *the most* by $\aleph(0)$ will have the greatest reductions in their perceived difficulty, and, according to our model above, will suffer the greatest reduction in status. Similar patterns occur in many instances of automation.

The centrality of questions, that is to say, their relationship to others, is another component of the way we value mathematics, and we expect $\aleph(0)$ to change this too. Let us suppose that the energies of $\aleph(0)$ are partly directed towards reworking the existing literature: revisiting and supplying proofs of known results rather than examining open questions. As we have emphasized, the number of mathematicians who have thought about a specific question is typically very small, and it is likely that very many parts of the literature would be greatly revised even through careful re-examination by many human mathematicians. It is not unlikely that we will see a scenario that has happened surprisingly rarely in recent history – replacement of long elaborate proofs by short overlooked ones. What effect might a five page combinatorial proof of the Weil conjectures have? Even if such an extreme scenario does not occur, it seems very likely that the web of relationships between standard lemmas and theorems will be altered. This discussion also suggests why the operators of $\aleph(0)$ may be induced to revisit old problems over studying new ones: besides settling concerns about formal correctness, the shifting of foundations has a larger social impact than adding new levels.

Finally, $\aleph(0)$ will greatly expand the entire landscape of questions considered mathematically interesting. Such inflation can happen through many different paths; it is not necessary for $\aleph(0)$ to explicitly generate questions on its own, for new mathematics always generates new questions, and correspondingly any process accelerating research in mathematics will accelerate the creation of new questions. (If $\aleph(0)$ does all the proving and we do all the questioning, the result is not so different to a scenario where $\aleph(0)$ is capable of generating its own mathematical conjectures.) Now we already saw that fields with an oversupply of problems relative to the number of workers may lose status, particularly if those problems do not organize around central ones. Since the existence of $\aleph(0)$ will increase both the number of problems and the effective number of workers it is not clear how this will play out; but certainly we may expect great variability from the current

situation. In such an expanded landscape, many currently central problems may become peripheral.

These three points already suggest a great shift in what problems and fields will attract the most attention. However, the process may extend beyond this, and affect, for example, the balance between heuristics and rigor, the role of aesthetic considerations, the extent of consensus, and the placement of boundaries such as those between professional and amateur mathematics or pure and applied mathematics. ($\aleph(0)$ will likely level the playing field between professional mathematicians and other interested parties.) To analyze the specifics is obviously impossible without a better idea of the abilities of $\aleph(0)$, but whatever direction it goes, it will go far.

An important limitation on rapid change in a subject is the the length of the professional career. Those who can most readily enter a new field are the young, and the extent to which this is possible is limited by the structure of hiring; senior scientists are slower to change their view of what is valuable. Nonetheless, since it will presumably be infeasible to do research without making use of mechanized assistants in the post-$\aleph(0)$ age, the impacts that we have detailed above will likely extend to senior mathematicians also, although their effects will be more extreme for younger mathematicians.

In the normal development of any scholarly field the way we assign importance and value is continuously changing and evolving. What distinguishes our scenario is the breadth and magnitude of these effects and the short timescale over which they are likely to occur; developments that previously took several mathematical generations may be compressed into a few short years.

It is natural to look to history for metaphors. Post-mechanization mathematics may look to us as modern mathematics might impress those working a century ago, but I think this does not go far enough. The impact of $\aleph(0)$ on mathematical cognition may be much greater than the passage of a hundred years. To find a suitable parallel for this effect on our thought process, we might consider, for example, the introduction of algebraic notation in mathematics.

It is important for us to consider seriously the possibility of such developments.

Acknowledgements. I would like to thank all the participants of the STMS seminar for providing a stimulating atmosphere in which to explore these ideas. I also thank Ken Alder, Aravind Asok, Brian Conrad, Harald Helfgott, David Nirenberg, Patrick Shafto, and David Treumann for interesting discussion and suggestions. That writing is a useful metaphor was suggested by Ken; Patrick pointed out the reference [5], and Brian pointed out several errors and suggested many interesting examples to consider. Finally, I note that the topics here have a substantial overlap with an excellent recent essay [2] of Jeremy Avigad which also includes many interesting historical and mathematical examples.

REFERENCES

[1] Abrams v. United States, 250 U.S. 616, 1919.
[2] J. Avigad, " Varieties of Mathematical Understanding," Bulletin of the American Mathematical Society, 2022.

[3] J. Pearl, "Reverend Bayes on inference engines: A distributed hierarchical approach," Proceedings of the Second National Conference on Artificial Intelligence, AAAI Press, Menlo Park, California, 1982.
[4] J.-M. Schlenker, The prestige and status of research fields within mathematics, https://arxiv.org/abs/2008.13244.
[5] S. Sloman, B. Love and W. Ahn, "Feature Centrality and Conceptual Coherence," Cognitive Science vol. 22 (2), 1998.
[6] H. Weyl, address at Princeton Bicentennial Conference, 1946.

Current revision date: February 21, 2022.