Logic and **L**ogical **P**hilosophy (2014)
DOI: 10.12775/LLP.2014.009
Published online: April 30, 2014

Jason Megill
Timothy Melvin

COMPUTABILITY AND HUMAN SYMBOLIC OUTPUT

Abstract. This paper concerns "human symbolic output," or strings of characters produced by humans in our various symbolic systems; e.g., sentences in a natural language, mathematical propositions, and so on. One can form a set that consists of all of the strings of characters that have been produced by at least one human up to any given moment in human history. We argue that at any particular moment in human history, even at moments in the distant future, this set is finite. But then, given fundamental results in recursion theory, the set will also be recursive, recursively enumerable, axiomatizable, and could be the output of a Turing machine. We then argue that it is impossible to produce a string of symbols that humans *could possibly produce* but no Turing machine could. Moreover, we show that any given string of symbols that we *could* produce could also be the output of a Turing machine. Our arguments have implications for Hilbert's sixth problem and the possibility of axiomatizing particular sciences, they undermine at least two distinct arguments against the possibility of Artificial Intelligence, and they entail that expert systems that are the equals of human experts are possible, and so at least one of the goals of Artificial Intelligence can be realized, at least in principle.

Keywords: Computability theory; Turing machines; axiomatization of sciences; artificial intelligence

1. Human symbolic output

Consider the various symbolic systems humans have produced; e.g., the natural languages such as English and Japanese; first-order logic; the symbolic languages used in different areas of mathematics, or music, or

Received August 5, 2013; *Revised* April 25, 2014

in the various sciences such as physics. In other words, think of any human symbolic system that has symbols and rules (or syntax) for how those symbols can be combined into well-formed or meaningful strings of characters. We produce these strings of characters constantly, e.g., any time we write down a grammatical sentence in a natural language, or a mathematical claim, or an equation in Quantum Mechanics. At any given moment in human history, there will be a set that consists of all and only those strings of characters in a symbolic language that at least one person has produced up to that moment. Call this set the "ASO," for "Actual Symbolic Output." Again, the ASO will contain every string of characters produced in any symbolic language up to a given moment in human history; so, for instance, at the present moment, the ASO will contain all strings of characters any one has ever produced in natural language, mathematics, physics, music and so on up until now. Of course, the cardinality of the ASO is continually growing larger with each string of characters produced by a human in a symbolic language; indeed, the cardinality of the ASO grows larger with each sentence we finish while writing this paper. And also note that not every string of characters in the ASO will correspond to a true claim; people often produce false claims in natural language, mathematics, etc. Some of these strings of characters might even lack a truth-value altogether. For example, the score of Bach's 5^{th} Brandenburg Concerto is a string of characters in a symbolic language, and so will be a member of the ASO; but, it doesn't appear to make sense to ask whether a string of musical symbols is true or false. Moreover, one can think of various subsets of the ASO; e.g., the ASO will contain the set of symbolic strings that have been produced in English, as well as the set of symbolic strings that have been produced in Japanese, or Quantum Mechanics, or Biochemistry, and so on.[1]

[1] There are some important and difficult questions involving the relationship between the symbol strings we produce and human thought itself. For example, can *all* human thoughts be encoded into a string of symbols or are some thoughts not (even potentially) symbolic? This question appears related to a different question: is thought necessarily linguistic or not? That is, must thoughts occur in a language? Also, one might wonder if there is a sort of universal symbolic language that underlies all of our particular symbolic systems, a sort of Mentalese (see [3])? These questions are important, but in strictly focusing upon only the strings of characters that humans output, we are trying to avoid them as much as possible. We can make the points we wish to make in the absence of answers to these questions.

Furthermore, one can posit a different set of character strings: the set of symbolic strings that humans *could* produce. Call this set the "*PSO*," for "Possible Symbolic Output." Of course, every string of characters in the *ASO* will also be in the *PSO*; if a human has actually produced a string of characters, then since we cannot do the impossible, it must be possible for us to produce that string of characters. But the *PSO* might very well contain strings of characters that we could produce but, for whatever reason, never will; e.g., perhaps there is some massively large addition problem that we are perfectly capable of producing, but will never bother to produce? And finally, one can posit a third set of character strings: the set of character strings that could be produced by a Turing machine. Call this set the "*TMO*," for "Turing machine output."

In the remainder of this paper, we argue that there are various relationships between the *ASO*, *PSO*, and *TMO*. In section two, we argue that the *ASO* is contained in the *TMO*; that is, at any given moment in human history, the strings of characters in symbolic languages produced by humans up to that moment could all be produced by a Turing machine, at least *in principle*. In other words, at least in *practice*, no human mathematician, physicist or poet will ever produce a string of characters that could not have been produced by a computer. In section three, we argue that the *PSO* is also contained in the *TMO*; that is, any string of characters in a symbolic system that *could be* produced by a human could also be produced by a Turing machine. We conclude by discussing some implications our claims have for attempts to axiomatize physics and for the possibility of Artificial Intelligence.[2]

2. Actual symbolic output

Consider the *ASO*. Again, the *ASO* consists of all character strings in any symbolic language that have been produced by at least one human

[2] The arguments in this paper build on arguments made in a previous paper. In [10], it is argued that any *mathematical* claim that humans will and could produce could be the output of a Turing machine. However, while the arguments in that paper focused *solely* on human mathematical output, here we attempt to generalize those claims to *any* symbolic system, including natural languages and the symbolic languages used in some sciences. This generates entirely novel implications involving, e.g., physics (as we discuss in section four). Moreover, some of the particular arguments given in this paper are not in the previous paper (e.g., the argument from mathematical induction to the claim that the *ASO* is finite that is given shortly).

up to a given moment in human history. We first argue that at *any specific* moment in human history, the *ASO* is (or was or will be) a set with a finite cardinality.

Note the elementary fact that (i) even though the natural numbers are countably infinite, any given natural number on the number line will be finite, be it 576, 48721, 91013144043 and so on. Also note that one can (ii) assign a natural number to every minute in human history in which at least one human has (or will) produce a string of characters in a symbolic language: label the first such minute "1," the second "2," and so on, perhaps to infinity (if humanity lasts that long). Each minute in human history will be assigned one and only one finite natural number. So, (iii) at any particular moment in human history, there will only have been a finite number of minutes — and so a finite amount of time — in which humans have been producing character strings. But then, since (iv) character strings take time to produce, (v) humans will have only produced a finite number of character strings at any given moment in human history.[3]

If one finds the previous argument questionable, for whatever reason, here is a second argument: (i) pick a random minute m in human history. For example, minute 1115678490, though it does not matter which minute we pick. Call this minute our "base case." Clearly, (ii) at this minute, there will have been (up to and including that minute) a finite number of people producing a finite number of symbolic strings. But also note that the same will be true of the next minute, i.e., of minute $1115678490 + 1$. So, (iii) at minute $1115678490 + 1$, there will also have been (up to and including that minute) only a finite number of people producing a finite number of symbolic strings. (iii) is true because in the minute that passes from 1115678490 to $1115678490 + 1$, there will be a finite number of people producing a finite number of symbolic strings; call the number of strings produced in that minute "n." The number of strings produced up to and including $1115678490+1$ will be the number of strings produced up to and including 1115678490 (a finite number) plus n (also a finite number), and so will be finite as well. Therefore, by mathematical induction, we can infer that, (iv) at any minute m through-

[3] The above argument seems unproblematic: (i) is a basic fact mathematical fact; (ii) is an innocuous numbering of minutes with natural numbers; (iii) follows from (i) and (ii); (iv) is a simple fact about the production of symbolic strings; and (v) follows from (iii) and (iv).

out human history, the total number of symbolic strings produced up to and including that point is finite.

We have seen that at any *particular* minute in human history, the set ASO of human symbolic strings that have been produced up to and including that minute is finite. The ASO will have members that are simply strings of characters; these characters might be letters from alphabets in natural languages, or characters used in mathematics, or symbols used in physics and so on. And again, this list — though always changing as people continue to produce new symbolic strings — will always be finite at any particular moment. This will be true even if human history stretches infinitely into the future: just as any particular natural number will be finite even though the number of natural numbers is countably infinite, the amount of time between the first human symbolic string produced and any *particular* (later) minute in human history will be finite, even if humans will produce symbolic strings for a countably infinite number of minutes. Furthermore, trivially, any particular human that will ever exist will exist in one or more of these particular minutes: there will never be a human that can "look backwards" at a countably infinite number of symbolic strings.

Given that at any particular minute in human history, the set ASO will be finite, then since finite sets are recursive, the ASO will be recursive too. So, for instance, one can — at least *in principle* — devise an algorithm that could determine whether or not some random string of characters is in the ASO (at any given moment). Consider some random string of characters. Check the first member of the ASO to see if this random string of characters and this string of characters in the ASO are the same. If the strings are the same, then output a 1 and stop. If they are not the same, then move on to the next member of the ASO; again, if the strings match, output a 1, and if they do not, then move on to the next member of the ASO to see if this member matches the string. One of two things will happen: we will eventually get a match (and so a value of "1"), or we will go to the end of the finite list without finding a match (in which case, output a value of "0"). We can take a certain input (a random string of characters), and then use an algorithm to determine if this string is or is not in the ASO. So the ASO is recursive.[4] And

[4] See, e.g., [14]. Basically, (where $n =$ a random string of characters), since there is an algorithm that returns the value $1 \leftrightarrow n \in ASO$ and the value $0 \leftrightarrow n \notin ASO$, the ASO is recursive.

note again that the *ASO* is a fairly interesting set. By definition, *all* of the symbolic strings that *any* human *will ever produce* will eventually be in the *ASO*; a symbolic string produced by a human at some point in human history will automatically become a member of the *ASO* when it is produced. By definition, we will never produce a symbolic string not in the *ASO*, because as soon as a string is produced, it automatically becomes a member of the *ASO*. So, human symbolic output at any particular moment in time, as we know it and as we will always know it, will always be *recursive*.

But if so, then a number of claims about actual human symbolic output are true. For example, since recursive sets are recursively enumerable, the sum total of human symbolic output at any particular moment throughout human history will always be recursively enumerable (see, e.g., Wang [14]). Furthermore, since recursively enumerable sets are axiomatizable (see [2]), actual human symbolic output will always be axiomatizable. Indeed, since this set will always be finite at any particular moment, it will be finitely axiomatizable. Moreover, the sum total of actual human symbolic output could be the output of a Turing machine. In other words, a Turing machine could produce as output any symbolic output that humans have ever and will ever produce.[5] The *ASO* is contained in the *TMO*.

3. Possible symbolic output

But now consider the set *PSO*, the set of all character strings that humans *could possibly* produce. Whereas the set *ASO* is finite at any particular moment in human history, the *PSO* is almost certainly (countably) infinite. A human *could* produce the character string "$1 + 1 = 2$," and "$2 + 1 = 3$," and "$3 + 1 = 4$," and so on *ad infinitum*. Nevertheless, the *PSO* still *might* be recursive, and so recursively enumerable, axiomatizable, and Turing computable; some infinite sets are recursive (e.g., the natural numbers, the prime numbers etc).

Whether or not the *PSO* is recursive, it is impossible for us to produce a specific string of characters that could not be the output of a Turing

[5] Note that a set is recursively enumerable if and only if a program exists that will give the members and only the members of the set as output (see, e.g., [14]). We have already seen that the *ASO* is recursively enumerable, so there must be a program for producing only the members of the *ASO*.

machine. Consider three sets. (1) The set *ASO* of symbolic strings humans actually will produce at some point in human history. (2) The set *PSO* of symbolic strings that humans could possibly produce. And (3) the set *TMO* of symbolic strings that a Turing machine could possibly produce. Above, we saw that a Turing machine could produce as output all of the strings in the *ASO* (as it exists at any given moment throughout human history); so the *ASO* is contained in the *TMO*. So, if there is a symbolic string that humans could possibly produce but no Turing machine could, or if there is a string of characters that is in the *PSO* but not the *TMO*, this string must be in the *PSO* though not in the *ASO*, for if it was in the *ASO*, then it would be in the *TMO* as well. But we will forever be unable to actually produce this string; for if we could produce it, then it would be in the *ASO* (since the *ASO* is the set of symbolic strings humans will produce), and so in the *TMO* as well (and so could not be used as a counterexample to mechanism). To reformulate the argument, suppose that we can produce a character string that humans can produce but no Turing machine could; indeed, we can produce this counterexample to mechanism. We could, e.g., write this string of characters on a chalkboard and proclaim mechanism dead. So, this string is in the *ASO*, and so also the *PSO*, but not the *TMO*. But we saw above that every character string that humans will ever actually produce could also be produced by a Turing machine; any string in the *ASO* is also in the *TMO*. So this string both is and is not in the *TMO*. The assumption that we can produce a character string that can falsify mechanism entails a contradiction; so we cannot produce a particular counterexample to mechanism. We cannot produce an example of a character string that could not be the output of a Turing machine.

Moreover, it appears that one can establish a stronger claim: one can show that any *particular* symbolic string that is in the *PSO* is also in the *TMO*. Consider the set *ASO* of actual human symbolic strings produced up to a given moment in human history and the set *PSO* of symbolic strings humans could possibly produce. As mentioned above, the *ASO* is a subset of the *PSO*. If a human has produced a given symbolic string at some point, and so this string is in the*ASO*, then it must be possible for a human to produce that string, and so this string is in the *PSO*. Furthermore, there will either be (a) at least one string that is in the *PSO* but not in the *ASO* or (b) not; this is an instance of excluded middle. If (b) not, then since the *ASO* is recursive (see above), the *PSO* will be recursive — and so computable — as well. But if (a), then the

PSO contains more strings than the *ASO*. Call the set of these strings that are in the *PSO* but not in the *ASO* the "*USO*," for "Unproduced Symbolic Output." The *USO* contains producible strings (again, they are in the *PSO*) that will — for whatever reason — forever remain unproduced. Now, consider a random string in the *USO*. It appears that this string *could* be in the *ASO*. After all, this string is possibly producible by definition (it is in the *PSO*), and the *ASO* is the set of those strings that we will produce; so clearly, this string at least *could be* in the *ASO*. But then any given individual string in the *USO* could have been in the *ASO*; with general conditional proof, if any random member of the *USO* could be in the *ASO*, then any given member of the *USO* could be in the *ASO*. But then any member of the *PSO* could be in the *ASO*, since the *PSO* will simply be the union of the *ASO* and the *USO*. And since a Turing machine can produce the set *ASO* as output (see above), it should be able to produce any string that is in the *PSO* as output as well. Any particular symbolic string that we *could* produce could be produced by a Turing machine. A Turing machine — not only in practice but even *in theory* — could produce any given human symbolic output.

4. Implications

We've argued that at any given moment in human history, the set *ASO* of symbolic strings that humans will have produced up to and including that moment is (i) finite, (ii) recursive, (iii) recursively enumerable, (iv) axiomatizable, and (v) could be the output of a Turing machine. Furthermore, the set *PSO* of symbolic strings that humans could possibly produce is such that (vi) we can never produce a specific member of the *PSO* that cannot be the output of a Turing machine and indeed, (vii) any specific member of the *PSO* is computable, at least insofar as any given symbolic string we could produce could be the output of a Turing machine. We conclude by briefly discussing some implications of these claims.

First, the arguments have implications for Hilbert's sixth problem, or at least a variant of it. Famously, in a lecture delivered at the International Congress of Mathematicians in Paris in 1900, the mathematician David Hilbert gave a list of 23 great unsolved problems in mathematics. One of these problems, the sixth, dealt with the axiomatization of certain areas of physics; Hilbert [6] wished "to treat. . . , by means of axioms, those physical sciences in which mathematics plays an important part."

Corry [1] claims that the sixth problem has often been seen as an odd inclusion in Hilbert's list of 23 problems; to many, the sixth problem doesn't appear to be like the other problems on the list.[6] Corry [1] claims that

> Thus, in reports occasionally written about the current state of research on the twenty-three problems, the special status of the sixth problem is readily visible: not only has it been difficult to decide to what extent the problem was actually solved (or not), but one gets the impression that, of all the problems on the list, this one received the least attention from mathematicians throughout the century and that relatively little effort was directed at solving it ([5] and [15]).

Nevertheless, if the argument in section two is correct, it does appear that areas of physics can be axiomatized, at least *in principle*. Consider the set of claims in a particular branch of physics — understood as strings of symbols in the relevant notation — that humans have produced at any given moment in human history. This set will be finite: it is a subset of a finite set (the ASO) and so must be finite itself. But if so, then it will be recursive, recursively enumerable and axiomatizable. That is, the set of claims in any branch of physics — as it exists at any given moment in human history, and so as it exists in the only way we will ever know it — can be axiomatized, at least in principle. And the same will hold for all sciences that employ symbolic systems.

Second, the arguments above have implications for the possibility of Artificial Intelligence. For instance, there are extant arguments that machine intelligence will forever fall short of human intelligence. The Lucas-Penrose argument is one example (see [7, 8, 9]; see also [11, 12]). The Lucas-Penrose argument posits a mathematical claim that we can see is true, a Gödel sentence (see [4]), but no Turing machine could; it posits a mathematical claim that we can produce but no Turing machine could. But clearly if a Turing machine can produce any symbolic claim that we can, we cannot produce a mathematical claim that a Turing machine could not.[7] But this point can be generalized to apply to strings produced in *any* symbolic system. For example, sometimes people claim that since computers "lack creativity," any machine intelligence will fall

[6] Though Corry himself rejects this view and argues that Hilbert's sixth problem is a natural outgrowth of Hilbert's work on and conception of axiomatic systems etc.

[7] See [11] for more on what our arguments imply about human mathematical output specifically.

short of human intelligence in at least one important respect. But if our arguments are correct, then any musical score, novel, poem, etc. that we can produce could be produced by a Turing machine as well. Musical scores, novels, etc. are simply strings of characters in symbolic languages, and a Turing machine can produce any string of characters that we will or even could produce.

Third, the arguments above provide some evidence that expert systems that are the equals of human experts are possible, and so at least one of the goals of Artificial Intelligence is realizable in principle. Of course, in A.I., an expert system is a machine intelligence that can reason like a human expert in one specific domain or area of discourse, whatever it might be. And if the arguments above are correct, it should be possible to construct expert systems that are just as competent as humans in a variety of areas; e.g., it is at least *theoretically possible* to produce an expert system in, say, mathematics, that could match Gauss theorem for theorem. Recall that the *ASO* is a finite set. The strings of characters that correspond to human output in a particular discipline, e.g. physics or mathematics, will be contained in the *ASO*, and so they will be finite in number as well; again, a subset of a finite set must be finite. But then all of these subsets that correspond to particular disciplines will be recursive, recursively enumerable, axiomatizable and Turing computable.[8]

References

[1] Corry, L., "On the origins of Hilbert's sixth problem: physics and the empiricist approach to axiomatization", pages 1697–1718 in *Proceedings of the International Congress of Mathematicians*, Madrid, Spain 2006. http://dx.doi.org/10.4171/022-3/82

[2] Craig, W., "On axiomatizability within a system", *The Journal of Symbolic Logic*, 18 (1953): 30–32. http://dx.doi.org/10.2307/2266324

[8] We resist the temptation to make stronger claims such as: (a) "our arguments show that Turing machines are the equal of the human mind," or (b) "our arguments show that Strong A.I. is possible" (Searle [13] defines "Strong A.I." as the claim that properly programmed machines literally have a mind, and so cognitive states, and an understanding of the meaning (or the semantics) of the symbols that it manipulates). Again, (a) it *might* be that some human thoughts cannot be converted into a symbolic language and so given as output in the form of a string of characters, and these thoughts are non-computable. Likewise, (b) there are arguments against the possibility of strong A.I., such as Searle's Chinese Room Argument (see [13]), that we have not addressed.

[3] Fodor, J., *The Language of Thought*, New York: Thomas Crowell 1975.
[4] Gödel, K, "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I",*Monash. Math. Phys.*, 38 (1931): 173–198.
[5] Gnedenko, J., "Zum sechsten Hilbertschen Problem", pages 144–147 in *Die Hilbertsche Probleme* (ed. by P. Alexandrov), Ostwalds Klassiker der exakten Wissenschaften, Leipzig 1979.
[6] Hilbert, D., "Mathematische Probleme", pages 253–297 in *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Math.-Phys. Klasse*, 1900.
[7] Lucas, J. R., "Minds, machines and Gödel", *Philosophy*, 36 (1961): 112–127.
[8] Lucas, J. R, "Mind, machines and Gödel: A retrospect", a paper read to the Turing Conference at Brighton on April 6th, 1990. http://users.ox.ac.uk/~jrlucas/Godel/brighton.html
[9] Lucas, J. R., "The Godelian argument: Turn over the page", BSPS conference, Oxford, 1996. http://users.ox.ac.uk/~jrlucas/Godel/turn.html
[10] Megill, J., T. Melvin, and A. Beal, "On some properties of humanly known and humanly knowable mathematics", *Axiomathes* (forthcoming).
[11] Penrose, R., *The Emperor's New Mind*, Oxford: Oxford University Press 1989.
[12] Penrose, R., *Shadows of the Mind*, Oxford: Oxford University Press 1994.
[13] Searle, J., "Minds, brains and programs", *Behavioral and Brain Sciences*, 3 (1980): 417–457. http://dx.doi.org/10.1017/S0140525X00005756
[14] Wang, H., *Popular Lectures on Mathematical Logic*, Mineolam NY: Dover 1981.
[15] Wightman, A. S., "Hilbert's sixth problem: Mathematical treatment of the axioms of physics", pages 147–240 in *Mathematical Developments Arising from Hilbert Problems* (ed. by F. E. Browder), Symposia in Pure Mathematics, Amer. Math. Soc., Providence, RI 1976.

Jason Megill
Carroll College
Department of Philosophy
Helena, Montana, USA
jlm3am@virginia.edu

Tim Melvin
Carroll College
Department of Mathematics
Helena, Montana, USA
tmelvin@carroll.edu