Translating the machine:

Rearticulating intelligence and cognition

Samuel Aronson

Senior Thesis

Department of Philosophy

Haverford College

April 22nd, 2022
Abstract

Advances in artificial intelligence technologies are displacing human labor and continue to gain new capacities. Given these displacements, particularly those that not just displace physical tasks but mental ones, there are fears that machines will be able to do all of what humans can, but better. Many accounts of mind flat-out reject the notion that properly programmed computers can have a mind but struggle to provide an account for what they can do. This thesis will sketch such an account. First, it will follow philosophers of mind in accounting for the mind as embodied and constitutive of many behaviors and capacities before unifying the mind and computational resources through an account of cognition. Finally, the topic of translating cognitive contexts will be considered. In all, this will allow us to radically reconceptualize how to think of not just mental occurrences, but the relation between people, machines, and other cognitive beings.
Acknowledgements

My most heartfelt of thanks to the Haverford Philosophy department; I can truly say that this department changed my life. My fellow senior majors are some of the most thoughtful and curious people, constantly reframing how I see the world; I am privileged to be able to call you my friends. The faculty have provided me the educational experience of a lifetime. Professor Yurdin: thank you for supporting me and providing me the structure as I embarked on my largest intellectual project to date. Professor Macbeth: thank you for constantly encouraging me to clarify not just my writing but my thinking. Professor Miller: thank you for constantly pushing me not just to write, but to risk something as I do so.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Minds, Brains, and Bodies</td>
<td>5</td>
</tr>
<tr>
<td>Reinterpreting the Chinese Room</td>
<td>9</td>
</tr>
<tr>
<td>Cognition, Thinking, and Knowledge</td>
<td>13</td>
</tr>
<tr>
<td>Translating cognition</td>
<td>20</td>
</tr>
<tr>
<td>Selected Bibliography</td>
<td>31</td>
</tr>
</tbody>
</table>
**Introduction**

There is a tremendous amount of excitement and fear about what advancements artificial intelligence are bringing and will bring to society. Already, it is aiding in the making of scientific discoveries, helping people drive more safely, and routing mail without need for human sorting. However, the capacities of technical systems continue to improve, and also are conducting tasks as hiring workers, driving, and writing news articles and other texts. With this comes an anxiety over what the role of so-called human or natural intelligence may pose for the future.

Automation and the displacement of human labor is not a new phenomenon. Like the march of automation that confronted such storied groups as the Luddites and brought forward an industrial revolution, perhaps artificial intelligence is the next chapter in this story. Is it simply the extension of the automation that has eliminated tasks, and thus jobs and livelihoods from manufacturing and other domains? However, in this case, the domain of automation is not that of physical but mental work; it is not work in the production of automobiles and dishwashers, but now the work of journalists, paralegals and scientists producing newspaper articles, relevant legal precedent and experiments. Often, it is believed that there are special capacities that the human mind holds, however it appears to be that machines can do at least some of this mental work.

Many people generally considered intelligent and often conducting work at the forefront of computing seem to express concern over the role artificial intelligence will play. In short, they fear that artificial intelligence poses a threat to “humanity,” be it the development of machines that surpass and replace the need for all human work, or at least to human independence. Stephen
Hawking parrots the latter concern, warning for artificial intelligence, “the long-term impact depends on whether it can be controlled at all.”¹

A second concern, raised by Elon Musk, among others, concerns the idea that artificial intelligences will become significantly more intelligent than the human intelligences it is supposedly supplanting, noting “As AI gets probably much smarter than humans, the relative intelligence ratio is probably similar to that between a person and a cat, maybe bigger.”² This of course raises a number of questions, like how would one quantify a comparison of intelligence (a cat, or even many ostensibly capable machines would likely score 0 on an IQ test, notes John Haugeland in Artificial Intelligence). Even other standard evaluations of intelligence, like the Turing test are qualitative, not quantitative in nature. However, this seems to be rooted in the understanding that intelligence is a single type of thing that itself is shared between humans and other beings (like computers) without distinction. Taking this concern seriously seems to stem from the idea that computers will not only be able to possess the same set of capacities that human knowers do but supersede them. Even Dr. Hawking’s concern about control seems to come from a similar anxiety, that with inferior cognitive capacities human autonomy will be usurped.

What is notable about these anxieties is that they explicitly concern what computers can do, not how they will come to do it. In that sense, the question of interest appears to be a purely functional one, abstracted from any of the processes by which one can come to exercise that same function. And yet, when pushed to the margins it is rare that two different processes

---

¹ Higgins, “Stephen Hawking’s Final Warning for Humanity: AI Is Coming for Us.”
² Piper, “Why Elon Musk Fears Artificial Intelligence.”
produce exactly the same result, and by definition, they are not done in the same way, thus changing some sort of secondary property about it. Consider the querying the python programming language the arithmetic expression $1.2 - 1$, which is returned as an unexpected and incorrect value of $0.19999999999999996$. While this is an extremely easy calculation to do in one’s head, the program produces an inexact approximation, because of the way in which decimals (floating-point numbers) are represented. The method by which the program attempts to perform this calculation fails at the edges, due to the way in which this functionality is implemented. Depending on the context, this approach may be more or less suitable. On the margins of any capacity there is likelihood for error. While I may consider myself proficient in how to multiply numbers together, there’s a decent likelihood that I’ll make a mistake calculating $589311 \times 958672$ by hand; in fact, I’d place greater trust in python’s conclusion that the product is $564955954992$, given the reliability of integer multiplication on computers. In addition, my computer can calculate that value thousands of times faster than I can, because it is, in that respect a better process. Without knowledge of when and how a process may find itself to be fallible or unable to complete a task, one cannot fully understand the capacity that is possessed, for it is a product, that is reliant on the process. Even further, it can structurally restrict the possibilities of what such a process can enable, allowing us to circumscribe its capacities.

Insofar as two processes are considered equivalent, it is also important to understand on what qualities equivalence is considered. My neighbor had a hen who laid eggs; in some sense, they had functional equivalence to the eggs that I bought in the grocery store. They could be used in the same recipes in the same manner. And yet there was a richness of flavor that came from my neighbor’s eggs that were due to the care that he put in his hens, the diet and life that the hen
was able to lead. This is all to say that both functionality and equivalence are both concepts that should be brought into question; they are also inseparable in some way from the process by which that function can be found.

To answer a question about the capacities of computers and technical systems, we thus need to understand the technologies that underly not just those computers, but also human thinkers and cognizers as well. This thesis will first consider accounts of the mind, including mistaken notions that allow people like Mr. Musk and Dr. Hawking to make such grave category mistakes that philosophers including Gilbert Ryle help dispel. This work will then follow John Searle’s argument that “artificial” intelligence is a categorically different thing as compared to human or natural intelligence, but explore the different processes that are now being pursued under the name of artificial intelligence. It is to examine in what ways these capacities are shared, and what helpful distinctions can be made when considering the evolution and histories of both human and mechanical capacities, loosely considered their “forms of intelligence.”

Following these distinctions, we shall return to the question of what unifies the cognitive capacities of both humans and non-humans. Following N. Katherine Hayles’ work with posthumanism, we will consider a broad conception of cognition, including acts conducted by humans, non-human forms of life as simple as bacteria, as well as mechanical and electronic technological systems. Then, I will describe the importance of the act of translating in understanding different forms of cognition, while attempting to avoid the ever so common anthropocentrism that typically arises when considering issues of cognition and mind. From this point, it shall be clear why so many common conceptions of mind are misguided and the way in which the contributions of artificial intelligence should be considered.
Finally, an account of how we can come to translate these acts of cognition among varying contexts is explored. It is from here that we can not only ask what computers may be able to do, and how they may displace current notions of work, but the other ways in which these technologies can displace and transform the human, beginning to ask an aspirational question as well that is not just “what will the machines do?,” but “what should humans do?” To begin, however, we must return to our understanding of the mind.

Minds, Brains, and Bodies

In the philosophical tradition, there is a long history of giving a separate priority to the mind. This confusion begins with Rene Descartes and the establishment of modern science. Shifting from the ancient conception that conceived of every thing as having a nature intrinsic to it by which it operates (for instance, fire’s nature was considered to rise whereas earth’s nature was thought to fall). Instead of conceiving things as having natures internal to themselves, Descartes conceived of the world as being subject to universal laws external to the being. However, insofar as the laws of physics were universal, and Descartes was a “religious and moral man” he could not accept that we are simply more complex mechanisms than, say, a clock, answerable only to the same laws of physics. To allow for physics but prevent the reduction of the human to physical forces, he separated body and mind, and the spaces in which they operated. He imagined the body as in space, subject to physical laws like any other physical body and the mind subject to its own separate set of external laws, maintaining such important concepts like freedom of the will. However, this move raises a large number of philosophical

---

3 Aristotle, “Physics.” I.4
4 Ryle, Concept of Mind, 11.
questions like, how does mind and body interact, and how might one reconcile both mental and bodily causes together (when the much simpler approach is to consider them together).

It also follows from this view that mental consciousness could then be separable from the body, given they are composed of different substances. Because of this conception, the idea of the possibility of the separability of the mind is still incredibly popular. This separability is what allows popular culture to imagine the possibility of “uploading” or transferring one’s consciousness fully, as is considered in The Matrix and Avatar. Perhaps, it is also what allows one to differentiate between “mental” and “physical” work. Whereas it has been commonplace for hundreds of years for physical work to be automated, the mental work doesn’t seem to be as common. Further, it may substantiate Mr. Musk’s claim, insofar as it appears as if machines have not just encroached on the physical capacities of humans, but also broken through the “firewall” to mental capacities as well, although there will be more to say on this matter. In any case, it should be clear that the mind should instead be construed as an embodied phenomenon, inseparable from the physical instantiation.

However, Gilbert Ryle argues that this account is not only mistaken but “absurd,” because “it represents the facts of mental life as if they belonged to one logical type or category (or range of types or categories) when they actually belong to another.” In this case, it is that original assumption that mental occurrences are governable by a second set of laws, instead of the same set of laws by which all other bodies are governed. By taking this dualism as a mistake, it is also clear to see why many may assume that mental work is impossible to automate, insofar as it is governed by a completely different set of laws than the laws that govern mechanisms. By

5 Ryle, 6.
dismissing this claim, we can begin to understand the types of mental labor that could be displaced by machines.

In response to a dualist viewpoint, one may be inclined to reduce the mind to the biological, normally in the form of studying the brain and the causes of brain activity. However, this too in isolation is a mistake. If we were to follow this line of reasoning, it follows that biology would fully determine not just mental capacities, but mental occurrences. Every word that I write on the page would have a cause fully reducible to the biological. Every word I know would be too. Consider the child born to native-born American parents in China whose first language is Chinese. It seems clear that nothing in their genetic or biologic makeup, despite coming from native-born English speakers determines the language they learn or necessarily their capacity to understand English without being engrossed in the culture. To argue that such a thing is biologically determined would be to have a view of biology that is so expansive that it covers all of human activity, despite having a domain that does not appear to explain or hint at an explanation of how different natural languages arise. Biological determinism would not only fully determine the languages I can learn, but the languages I do learn, even though that appears to be a much more complex social phenomenon.

Instead, one should consider that many of the capacities that we associate with minded beings, such as being able to express one's thoughts and think in a language; being able to reason about concepts and objects in the world require notions that are embedded culturally, that is, beyond biologically. This is to say that to have a mind requires an upbringing that is in some sense social, in relation and interaction with others insofar as we understand the mind.

This is not to say that one’s biology or even the brain isn’t important or necessary for human thought, just that it is not the sole determiner, sufficient to provide an account of the rich
cognitive experiences that one credits with humans. The brain is essential to the workings of the mind; this statement is so obvious it barely requires stating, and impacts to it can affect one’s capacities for thought. Strikes to the head can have many impacts on mental occurrences and capacities like mood, memory, and one’s cognitive abilities. This is also to say that the ways in which the mind and brain is formed and developed is of the utmost interest.

Some biological determinists contend that human behavior is determined in its totality by biological (read: physical and genetic) makeup of its subjects. However, this ignores, as Catherine Malabou argues, that the brain too not just has a history marked by the process of evolution but “is itself a history,” a product of one’s own particular development. A deterministic view imposes a rigidity on the brain and thus often the mind, thought to be equivalent, displacing possibility from the human to the scientific and genetic. However, neuroscientist Jean-Pierre Cangeux emphasizes the plasticity of the brain not just as a property that allows it to be formed, but to form: “The error is in thinking that neuronal man is simply a neuronal given and not also a political and ideological construction (including of the ‘neuronal’ itself).” This is to say that the possibilities of the brain are expanded, its capacities contingent on its history: its practices and the conditioning of itself over time, the occurrences that have struck the organ from physical and psychological sources. This expands even the biology of the brain, as one constitutive (but not all-encompassing) element that develops the mind to having the capacity to shape shift, to reconstitute itself in a dynamic and everchanging manner. This is to say that any human cognition, any human intelligence, even that rooted in the brain itself is

---

6 Malabou, *What Should We Do with Our Brain?*, 1.
7 Malabou, 13.
inextricably linked with its history. This is a condition that the scientific enterprise is ill-equipped to handle, given it imagines its objects of study as ahistorical, attempting to construe it with only the relevant history to the scientific question. On this account, such a history could never be fully bracketed, encoding instead a possibility not purely encoded in its origin, but evolving and growing into its potential in the Hegelian sense.

Finally, we can acknowledge through its concept acquisition that the mind can be self-conscious of itself in an important way, recognizing its memories and histories as its own. These are the structures that allows it to function and possess the mental capacities that it has in the ways in which it has it, such as interpretation. It also allows for a notion of the self that can evolve over time. It is through this self-recognition that can allow for important episodes of self-reflection and self-criticism. Through this even one’s core beliefs can be challenged, in a way that not all other beings may be able to do.

At this point, a sketch of a necessary account of the mind has been made. Following Ryle, the mind is always embodied, and is never separable from it. The mind is embedded in a cultural context, which provides it conceptual capacities that it can use upon acquisition; it is in fact a history that represents the manner in which it has come to form, that form being biological. Self-consciousness and the features associated with it were also highlighted. Now, the question must be addressed as to whether computers possess the same capacities and structures.

**Reinterpreting the Chinese Room**

Fifty years ago, a movement in artificial intelligence research developed to claim that through proper programming of formal symbol manipulation human intelligence could be
attained, rooted not specifically in the brain but in its functional capacities. Typically, computers are thought to fundamentally compute through the model of a Turing Machine, a machine that has a defined set of states and a storage “tape” as well as a function that takes states and can manipulate both state and storage given their current statuses. While no modern program is written by constructing and implementing a Turing Machine explicitly (that is, unmediated through anything but its formal mathematical description or a direct implementation thereof), nor do computers often have a physical tape, the Turing Machine is still the model of computation that all computer programs can be theoretically reduced to, establishing what can or cannot be computed. Instead, to briefly describe the current process, a program is written that is then translated using an interpreter or compiler into machine languages which then produces the specific instructions for a machine that also theoretically reduce to a Turing machine, although in practice they do not. This is to say that a computer is understood to have the capacity to manipulate symbols and encoded data in preconfigured ways. While those are not instructions for a Turing machine precisely, it is still computationally reducible. The hope was, as John Searle presents it that “the appropriately programmed computer really is a mind, in the sense that computers given the right programs can be literally said to understand and have other cognitive states.”

If intelligence was mere manipulation of symbols in particular forms that this would have bode well for this project. However, as Searle, Hubert Dreyfus, and John Haugland among others are keen to point out- such a form of intelligence is severely lacking. This need not be because

8 Searle, “Minds, Brains and Programs,” 67.
of the embodiment objection already raised but because cognition is not entirely circumscribed by such a formalism. For Searle, this objection is raised through the proposition of the famous Chinese Room experiment. Suppose that a human who does not know Chinese is placed into a sealed box that has a printer that spits out Chinese characters, a guide that is comprehensible to the person on how to respond through Chinese marks that can then be returned through an output slot. Searle argues that because the person can never interpret or understand the Chinese, that even though they may be presenting the appearance of knowing Chinese given they can navigate the scenario, they are not in fact understanding Chinese because it is remaining uninterpreted. While saying it remains completely uninterpreted may be a step to far, it is certainly the case that the language is not being interpreted as a person who could read Chinese would, and that with that comes additional capabilities that are important to our conception of understanding. Perhaps this is due to the fact that the context in which a computer operates is radically different than that of a mind, given the shortcomings of the symbolic environment that the computer is in. In some sense, this bears the difference between the Chinese translator locked in the room, and a translator out in the world. We can mark this type of paradigm, by which computer programs are the obvious analogy, by recognizing that it is not embodied or situated in the world in the same manner that the mind is, responsive only in the ways in which it has been programmed. This marks an extraordinary difference compared to a person’s standard cognitive capacities.

The room operator’s faculties of the language cannot be applied in the contexts one would expect them to. Note that, even for human language learners, these contexts can vary widely. An academic learning a second language to read primary texts may not have any faculties to speak or listen to the language (the same types of disabilities may occur in the deaf or hard of hearing as well, or alternatively the blind person who could not read written text). This is
to also say that even within a natural language such as English, there may be specialized vocabularies that makes the written word nearly unintelligible to a native speaker (try attending even undergraduate thesis presentations in an unknown discipline- in this case the exposure to the field will be relatively limited still). Note that even now, we often contort ourselves to be intelligible to others, including machines. Be it trying to order in the native language of a foreign country, or more interestingly, the way in which a student fills a Scantron test. In this case, special care is taken to make their answers readable by the machine, responding to a multiple-choice test (already an unintuitive way of configuring a question without the presence of this technology) and carefully bubbling the assumed answer. It is through this method that the limited interpretation that is available to the Scantron technology to be properly conducted. Again, as in the Chinese room example, the Scantron is unable to evaluate any student’s aptitude or claims about the domain examined on the test except in exceedingly specific circumstances, making its capacities real (consider the efficiencies it brings to standardized testing and large classes), but exceedingly limited. Searle argues that this tool can only produce knowledge in so far as it is an extension of a human user’s intentionality. In cases like these, where it is clearly configured by a human such an account seems to hold some weight. However, even without such a consideration of intentionality the realm of understanding could only even be considered exceedingly limited, if not inexistent. One would consider someone with Searle and others argued, correctly, that because the mind does more than mere formal symbol manipulation such a conception of an artificial human-like intelligence is in and of itself lacking. This distinction of artificial intelligence has become more apparent through the empirical research that has been conducted, meaning outside of specialized domains this type of symbolic artificial intelligence research has sputtered due to its unviability.
Instead of “good old-fashioned AI” approaches such as this one, current artificial intelligence approaches almost uniformly involve at least some machine learning techniques, where the computer is trained not given specific rules, but typically data and a method of learning from the provided data. In some cases, the machine is constantly receiving new information and changing its response due to it. In these cases, the machine appears to be “out in the world,” and in a relationship with the world in a different sense than the explicitly programmed computer. In the explicitly programmed computer, or the Chinese room, it is not responsive in any respect except the ways in which it has been programmed to be.

The differences between human minds and computers has been established and by reinterpreting Searle’s account some of the questions have arisen about how computers interact with what is widely understood to be the world today. Following Searle, we can still recognize the limitations of how these machines are in the world. Now, we must ask how we are to account for the similarities between what machines, humans, and perhaps other forms of life can do in relation to mental occurrences.

Cognition, Thinking, and Knowledge

Since the time of antiquity, rationality has been a capacity often reserved for human beings, likely excluding other known forms of life like animals or plants. Aristotle began by defining human beings as rational animals, with rationality being the exclusive domain of human soul that separate them from the souls of other living things. While there is nothing necessarily

wrong with defining a term in any certain manner, such as that which makes human beings
distinctive, one ought to consider whether it is a useful distinction to draw, or if such a property
is even ascertainable. In what ways can we account for the capacities that are shared between
humans and non-humans?

There are many conceptions of knowledge throughout the discipline of Philosophy. Many
accounts will either implicitly or explicitly refuse to credit non-human agents with knowledge,
categorically refusing things like other animals or computers as possessors of epistemic
knowledge. It may be because knowledge is typically concerned with high-level mental
operations often involving abstract thought, the use of language and other concepts that may be
involved in doing things like proving mathematical theorems or composing music. While these
capacities involve some of what it means to think and act, there are also occurrences credited as
mental that may not require a directed consciousness in the same way. This is possible again
because following Ryle we can recognize that mental capacities are not ontologically separate
from other capacities that a human may have. Consider the tennis player who makes a brilliant,
strategically placed shot but who must react so quickly that they don’t have time to think; or the
driver who is zoned out but is still able to drive competently, even braking in unexpected
emergency situations. In these types of situations, while still ascribable as a mental reaction, it
does not seem to be a conscious process, at least in comparison to writing a paper or engaging in
a substantive conversation. N. Katherine Hayles suggests these types of unconscious mental
occurrences could be shared with other living beings and technical systems, and thus be the layer

11 As an example, consider the treatment of computers in Green, “Evaluating Distributed
Cognition.”
by which we can come to understand the shared capacities among humans and the non-human systems of interest. It is also important an account such as this one to consider relations with non-human entities; while granting them knowledge may be controversial at least on some accounts, it seems like cognizers may be an appropriate term.

Such an account of cognition has been developed by N. Katherine Hayles, who proposes an account that unifies the lower-level cognitive abilities of humans with other beings, bridging processes across contexts and mediums of cognition. Her account does not focus on the materiality or physicality of the process as those are wildly different across bacteria, computers, and humans. Instead, following fields such as neuroscience she focuses on their functional as well as structural similarities:

Notwithstanding the profound differences in contexts, nonconscious cognitions in biological systems and in technical systems can share certain *structural* and *functional* similarities [to nonconscious human cognitions], specifically in building up layers of interactions from low-level choices, and consequently very simple cognitions, to higher cognitions and interpretations.  

There are a number of things to note about the similarities of these phenomenon, all of which will eventually be grouped under the term “cognition.” First, Hayles is cautious to not just mark out functional similarities (that is, what can be done). As already established, we cannot think of just what can be done but how it is, and Hayles marks out structural similarities in how these cognitions come to be, despite being realized in different contexts and in somewhat different

manners. This “layering” approach to higher level cognitions can help one understand how different beings can reach similar types of capacities.

Perhaps this can also help us understand how intuitive notions of a phenomenon can be established. It is of course notable that it is not just humans that can develop intuitions about a phenomenon, but also other beings too. For may dogs, a car-ride can cause them to be nervous, anticipating an unpleasant trip to a veterinarian. In the same way, one can understand the goals of many artificial intelligence programs to be attempting to develop similar intuitions for the machine as intuitions are developed for a tennis player or a dog; training it to differentiate between various phenomenon given a certain set of inputs and expected responses. In all of these senses, we can imagine a cognitive relation between the cognizers and the “world,” or the external context from which it is developing its intuitions or cognitions. Insofar as this is a viable model, we ought to have clarity on what cognitions in this sense are. Note that already, even advanced tasks like driving seem not to necessarily need to involve thinking but possibly mere cognition, opening it up for the possibility that computers will be able to do these types of tasks. Automating driving would cause again massive labor displacements, but marks the advanced possibilities that are possible for machines under this account, while still restricting some of the higher level self-conscious thinking that may not be possible for machines to do. This is all thanks to considering how the machines will be situated.

Hayles proposes a definition of cognition that can be used to bridge the divide between cognition in systems and diverse forms of life: “Cognition is a process that interprets information within contexts that connect it with meaning.”\textsuperscript{13} The implications of this definition can help

\textsuperscript{13} Hayles, 792.
construct a conception that can delineate important differences not only in the context of mind and artificial intelligence, but even between different types of beings. First, this definition must be parsed.

_Cognition being a process_ transforms it into an act as opposed to an attribute or even knowledge that one possesses. This helps separate it from any static form of knowledge, like words on a page into a dynamic act, even if that is simply a person reading those words. Only in action does cognition occur. Importantly, cognition doesn’t necessarily require any level of self-consciousness, like the level which humans possess but likely not animals and certainly not plants. It also allows computers and other technical systems (even ones as simple as an analog thermostat) to cognize but again does not require self-cognition, preventing the necessity of the hysteria of the self-conscious “AI overlord” despite advanced capacities. It allows delineation between the type of the thing that can cognize and has what is generally considered human-like autonomy. The unity of the act of cognition is essential to the clarity that this concept brings. In this process it is also to say that the act of cognition is contingent on the faculties of the medium of cognition. The medium of the physical body, including the brain, is different in its constitution than the computer and the possibilities of its sensory capacities. The differences are in part constitutive of the context in which cognition occurs. This also opens for any active process to potentially be an instance of cognition so long as there is a response to something that could be considered information.

_Interpreting information within contexts_ allows cognizers to recognize the radically differing environments in which information is processed. This recognizes that any signal, be it words on a page, a birdsong, or machine code can be interpreted but that they are all within a specific and differing context. In this case, one is defining information to be a fusing of
information with that which has already been situated in the environment. It also allows for the context of the cognizers to have varying levels of depth and complexity, even recognition of one’s own history. It also provides a method of visibility into disability that can often be overlooked when simply considering a standard human or standard cognizer. This allows for the fact that even for certain human beings (actually, to some extent every human being) their context is different and may be constrained by various physical or biological or ideological constraints to how their information processes may occur.

It is also worth noting that this definition of interpretation conflicts with John Searle’s claim that the participant, human or otherwise, in the Chinese Room experiment is not interpreting anything. In fact, it is clear that cognition is occurring in this account, a fusing of information (the incoming characters) with context (the procedures) in an act. While computers on this account would be credited with interpretation, it would be a radically different context than the type of interpretation that Searle imagines a fluent Chinese speaker would be conducting. It goes without saying that their context is much richer and wider than the person in the room. Such an account is consistent with the critique that Margret Boden raises in response to Searle. Following from ideas in logic, Boden presents the idea that in the context of a computer program there are two different meanings of a variable, the first as the type of symbol that is manipulated internally by the computer system and the second as “a variable in logic [that] is usually some object external to the formalism itself.”

Through this program, we can be in a position to have confidence that some set of procedures can tell us something about the outside world, not because of what is internal to the program, but contextualized we can

---

understand the procedure to be applied. This allows for secondary understanding of these cognitions as well.

Of course, of particular interest is how such contexts are formed to begin with. It too seems to be iterative upon one’s experiences and feedback loops of those sources of information, sensory or otherwise. While every body will have different contexts, no one ever being exactly the same as those experiences are different, there are simultaneously communities and generally shared contexts that are also helpful in understanding such intersubjectivity across differing types of beings.

The available context is also, like in the case of the mind, a result of its available faculties; the waves of radio communication are simply unavailable for cognition to human beings in a direct manner because they do not have the sensory capacities required, while they certainly are for a cellular phone. This also emphasizes the importance of medium, form, and the possibility of translating between contexts, like a cell phone can translate between the auditory signal produced by a speaker and the radio waves that transmit that same signal, rendered inaccessible directly to that same signal producer.

Finally, Hayles argues that the cognitive process connects with meaning. While meaning might be construed as something fixed, it is clear it arises from this fusing of signal and context and is thus dynamic as new context arises with the introduction of a new signal, and the medium hosting this context, like the brain, can be evolving and changing as well over time. This also begins to conceptualize meaning perhaps as the map between these connected occurrences of information and interpretation. Finally, of importance to Hayles’s work but less so in this project, such an account places the nonconscious cognizer as legitimated subject of study in the humanities insofar as the humanities are understood as the pursuit and study of meaning. At this
point, we have established cognition as an interpretive process that is both dependent on the context or situation as well as the structures by which cognitions are developed. We must also understand how it cognition can be used and understood amongst disparate instantiations, and for that the question of translation shall be undertaken.

**Translating cognition**

In some recent work in epistemology, there has been increased focus on the social aspects of knowledge acquisition, typically between human agents. This seems to stem from the intuition that in many cases, an epistemic agent relies upon the claims of others to make their own claims, often in ways that are not confirmed by the agent. Joseph Shieber has document how even in the 1800s, cognitive work was being distributed similar to the way in which physical labor was being distributed and regulated during the industrial revolution.\(^\text{15}\) He begins with a discussion of Charles Babbage, who in turn describes de Prony’s project of producing mathematical tables of logarithms and trigonometric values for Napoleon. This was a vast project included different types of nearly 100 people working together, doing three different sets of tasks in a hierarchical manner. In the first tier, there are those who directed the project. They derive helpful formulae that ascertain the relevant information, but those persons play no role in the execution of the calculation of explicit instantiations of any formula or in the tabulation. In the second tier, there are a number of people with some mathematical ability that convert the formulae into specific arithmetic calculations to be executed. They then distribute these to be calculated by the third

\(^{15}\) Shieber, “Toward a Truly Social Epistemology.”
tier, who know little mathematics beyond simple addition and subtraction. After calculation, they are returned to the second tier to verify correctness (not by re-executing the calculations, but by checking its consistency with other results). Eventually, all results are tabulated, and the project was completed.

Without a doubt this is an instance of social cognition between human actors in this instance, it should be somewhat obvious that at this point the type of work conducted by the third tier of workers could easily be conducted by non-human cognizers, namely a calculator or general-purpose computer. In fact, it would feel most natural to do such a calculation using a technical system, if not the second “confirmation” tier as well. This should establish that like physical divisions of labor, mental or cognitive divisions of labor are also not a new phenomenon. It is also to mark that it is possible to divide cognitive labor between human and non-human cognizers, although of course this was not done in the nineteenth century. Thus, a model for how man and machine are able to work together can be sketched. That is to say, there are organizational modes for the division of cognitive work, just like there can be for physical work too. In the same ways, these can be split between workers, and between humans and non-humans.

If we are to engage in these cognitive relationships, the contexts for the various cognizers may be different. Consider again all of the instances in which cognitive relation between human and non-human is established. Already, technologies like the Scantron test have been used to examine students. Even in the case of a calculator or a computer, new ways of interfacing with the machine must be learned. Instead of writing an arithmetic expression or line of text, it has to be keyed in a particular manner. Typing is a skill that must be developed and is developed specifically so that one can type on a computer, making one’s words and actions intelligible to
the machines for cognition. While what would be considered sensory input is markedly different, these bridges can clearly be crossed.

Generally speaking, the act of acculturation to enable translation occurs whenever one is learning a new context, which need not be a new language necessarily as considered of the differences from, say English and Chinese or French. Consider the beginning Philosophy student who may not be learning new vocabulary per say, but instead is introduced to that vocabulary in a new light finding the import of a distinction, term, or rhetorical argument. The same can be found when learning mathematics: what constitutes a legal operation or step of justification of a proof must be learned. It is in many ways unintuitive in an important manner. These do not arise naturally but require the disciplining that occurs when being brought into a discipline, or really any context or language. In this sense, machines can be considered another context that must be learned, although the capacities that a machine may possess and the processes by which they possess them could look quite different.

Many mediums are inaccessible to all human observers, and demand technical systems or other life forms to mediate and act as a translator. Consider the K-9 dog that sniffs for drugs and explosives, or even the telescope that allows for human observers to see phenomenon in the galaxies inaccessible to the naked eye. These are allowing translations of objects that are deeply relevant to human inquiry yet are conducting translation impossible for humans to do alone. Instead, we have capacities of higher-level thought, the abilities to reason, to understand phenomenon like infrared light which may be inaccessible to human sensory experience directly. We can use technical systems and other tools to make such things accessible, and to understand the contexts in which other systems and forms of life operate.
One can consider the discipline of Computer Science to be doing enabling many of these types of translations. In addition to having its own disciplinary language, it also acculturates its students into studying the meaning embedded in computer systems. Computer Science acts as a translator to computerized systems from the cognition of its students. This study of translation occurs at many levels: everything from the high-level programming languages and expressive tools such as Python and Java which then are translated to machine code, low-level instructions which eventually get translated to the binary instructions upon which the exceedingly complex computers operate. Through the learning of these processes, one can translate requests relevant to human users to instructions that can bring relevant meaning.

In addition to the expressiveness of such programming languages, many aspects of computer programs can be formally proven by translating their constraints into mathematical statements, that is, a series of axioms about how an idealized computer operates that allows us to set theoretical constraints on what can even be evaluated. For example, it is provably true that a computer program that returns all prime numbers will never terminate. It is through this act of translation that we can prove things about mechanisms that insert themselves into everyday operation, guaranteeing in some cases results and expected performance. It is through translation of mathematics to such a domain that this understanding can occur. Of course, we also translate our own thoughts for computational compression constantly, as has already been elucidated by considering cases like those pertaining to the Scantron.

I wish to return to Margret Boden’s objection to John Searle’s Chinese Room experiment which claimed that the formal symbol manipulation that is occurring in the Chinese room has meaning, by virtue of the instructions that it conducts, even if it is not recognized by the cognitive acts of the machine. In many ways, this account is consistent with the Haylesian
account of cognition that this paper promotes. However, this articulation should not be considered as the human programmers impressing form onto the computer and thus giving it meaning; this is an interpenetrative act. Like Catherine Malabou argues for the brain, I wish to argue that a computer has plasticity, the ability to give and receive form. Like before, the medium gives certain advantages (for instance, doing arithmetic thousands of times faster than any human can), and other disadvantages (the computer’s memory of its past is much more tenuous than in most humans). It is through this that one can translate between human and computer meaning, using many technologies as tools to mediate the translation. This is to say that we can think of the relationship between human-based knowledge and cognition of machines as also occurring through this acculturation to translate between contexts. That is, there are things that the structure of the machines we have developed that are circumscribed as possible or impossible, easy or computationally complex to calculate. This is due to the differing nature of what computers are and how they operate from humans.

One can even consider this when conducting biological studies on even simple life forms. Bacteria’s environment, their context, is markedly different than that of a scientific researcher. One can consider the task of the researcher to learn the context in which bacteria cognize, to be able to place themselves in some way into the relevant conditions that can be made to reproduce the phenomenon of interest, despite this cognition not possessing the self-consciousness that can be attributed to the mind. That is, even this scientific researcher’s task is to translate between a scientific context and the context of the object of their study.

Of course, the topic of translation leaves one obvious topic: that of automated translation services like Google Translate which proprot to translate between natural languages like English and French. While it should be clear that this paper does not dispute the fact that some form of
translation is occurring, it is an inferior service. It translates websites in an unclear fashion, and any reputable piece of text published today will be translated by a professional, usually an expert in the domain in which it is being published. While the service is in some ways quite good, one can observe that it certainly does not do a perfect job of translation, or even one comparable to a human translator fluent in both languages. It should be clear that these acts of translation are not the same, and the Google Translate, instantiated using its current techniques will not be able to do the same thing. Currently, Google Translate uses advanced machine learning models on already-translated texts to learn how to translate between languages. That is, it is dependent on the work of previous translators, and the skill of human translation, to do its work because it is improperly situated to acquire a language, unable to learn a language in the same manner as a person.

Human translators do not usually learn how to translate simply by translating phrases and sentences, usually not documents either but by immersing themselves in the relevant context, which is of course dependent on what is trying to be done. It allows immersion in the intractably rich, embodied traditions associated with language; traditions people are often a part of in visceral ways. Imagine the connotations between traditions like grilling, or studying for exams, or steak-frites. These move past merely translating the word and become phenomenon, histories, and memories of experiences both personal and communal. Google translate is not embodied in a way to directly experience or understand these concepts, but instead must indirectly intuit them through textual artifacts. The mode in which it is immersed is different and will thus lead to differing translations, particularly on the margins of the translation, where judgement and
understanding seem to play an important role, instead of the mimicry that the automated translator is reduced to.\textsuperscript{16}

It is in fact the memory of histories that is an important difference that should be highlighted. A machine’s memory is not the same as a person’s. Its memory and practices can be wiped and reset on command. While technical systems can be fantastically complex, they do not have an understanding of such complexity in a self-conscious manner. They are certainly unaware of the context into which they are being brought into. Such acculturation processes look markedly different than how a person is typically brought into the world. This leads to a different shape of being, not hierarchically or universally but contingently in this moment. This complexity should and is also admired, but is different, and points towards the differences in capacities one can find between human, machine and other nonconscious cognizers. Such a history should constantly be appreciated, but as a similar but distinct enterprise from the environments in which humans cognize.

I hope to return to the concern parroted by Dr. Hawking, that is, the fear that as opposed to us controlling artificial intelligence that it is artificial intelligence that will “control” humanity. In many ways that world is already here. However, note that such translations between human and what one may call a limited machine intelligibility, like that which occurs when bubbling a Scantron form, are not only possible but quite common. While they can be in many cases

\textsuperscript{16} The question of the mimicry may have more than meets the eye. On one account, the mimicry is simply an incomplete simulation of an original phenomenon, perhaps missing some important element of how the task is done (for instance, the difference of embodiment mentioned here). However, it seems as if every language acquirer begins by mimicry. Where is the point at which the mimic becomes the authentic or the authorized? It does not appear as if there is always a simple or clear answer.
apparent and quite useful, like typing an arithmetic expression in a calculator to find the result, they can also have profound impacts like the automated applicant trackers used in many modern corporations. This type of software manages the hiring process for most large corporations. A feature of these systems is that they often conduct the first pass, reviewing an applicant’s resume determining whether they are a viable candidate worthy of an interview. In response, common advice I received when applying to internships and jobs was to modify my resume, so it became understandable by the software reviewers. I changed the descriptions of my previous experience, as well as added a section of just buzzwords of technologies I had a tight (or potentially somewhat lose) grasp on. One person I know even added this section in white text, unreadable to any human reviewer. My resume was translated to conform to the norms and expectations of the machine making these evaluations. In many ways, this seems to be an actualization of Dr. Hawking’s concern that the computers will be controlling us. And yet, this software would not be considered to be an “agent” in most of any sense, just simply a tool being used by these corporations in a somewhat mundane manner. These tools too at best have limited capacities of understanding, demonstrating that these fears need not be tied to intelligence as its understood qua human.

However, this control is not a novel phenomenon, or even one that requires the type of advanced AI that Hawking seems to be imagining. The contorting of oneself for the sake of a resume did not start when applicant tracking systems arose. Doing something for the sake of putting it on one’s resume is not a new phenomenon. This disciplining process starts whenever one is making themselves legible or attractive to a system, mechanical or not. During a job interview (the only one I attended for my future employer), I was interviewed by an Indian subcontractor, where I spent ten minutes explaining what a summer camp was as a part of my
professional background because that was not a part of his cultural context. In the same sense, the representation of myself needed to be translated to this quantifiable, objectified ideal and needed to be made intelligible.

Consider the demands of a form at the DMV, be it declaring one’s gender or hair color, height or weight. While these types of statistics and tabularizations are a tremendously good fit to the type of tasks that computers can do well and efficiently, with the advent of such categories, even before computers were on the scene, it is the advent of these categories that make such disciplinary power a method of control. This is all to say that such control has already been converted, and should not be a new anxiety, not that it shouldn’t be a force that is considered.

And yet, we continue to secede control of various decision-making processes to these technical systems, ones with significant societal impact. What is notable is that it appears that what is being developed is again sets of intuitions. It’s exposure to the world is through a very specific textual context. These machines are unable to manipulate in a general matter or make appeals to higher level principles that require thought and reflection like justice or fairness except insofar as they are predefined and inscribed. In some senses, it inscribes a static nature to the mode in which it can navigate the world if such navigation requires higher-level thought.

In recent memory, the response to the continued advances of automation, in a maneuver to displace the responsibility of job losses to the mis-trained worker, the instantiation of the college degree and white collar or “knowledge” work as the ticket to stable careers safe from the displacement that automation has conducted for labor. On account of this thesis, it should be clear that such a clean delineation between the safety and stability of mental as opposed to physical work is a mistake. It should also be clear given the advances of automation in many
types of knowledge work, such as that of the paralegal or newspaper journalist are at least in some of the tasks they do quite immense. For example, large swaths of the Associated Press’s stories regarding corporate earnings have been written by computers for years.\(^\text{17}\)

Simultaneously, many simple forms of “physical” work remain uncharted territory for automation: we are still in desperate need for plumbers and electricians, for fruit pickers and package stowers despite these being in many ways quite “simple” physical tasks. This is to say that the medium of the human body is still the most effective for much of the work that is to be done. This is also to say that the delineation of automation as occurring at the mind/body distinction, like that dualism in the first place, is a huge mistake. So, we can be clear that computers are not coming for all of human cognitive work, certainly not at the same pace because, following Searle, human and machine cognition are, at least as human is normally construed, different types of things.

If technology is really coming for your work: writing articles and emails as well or better than you can, make your food and do it as well as the humans doing those jobs, then it may be time to reconsider the nature of the work being done. Typically, machines are best at doing work that is regulated and regular, uncritical at the object of its work. Is that the type of writing one hopes to produce? If the practice doesn’t have value to the practitioner and the product is the same or better, why is that the quality that is being aspired for? Accurate and efficient reporting of financial statements is good, but is that what great journalism looks like? Is that what a great or piece of writing looks like? These mechanical tasks are important, and sometimes even importantly to be done by people to learn (after all, to write well one seems to need to write

\(^\text{17}\) Miller, “AP’s ‘robot Journalists’ Are Writing Their Own Stories Now.”
poorly first). Why are we spending time contorting ourselves in ways that are so easily mimicked by machines? N. Katherine Hayles notes that in the most recent movement for artificial intelligence, instead of machines being developed to meet human skill, it was instead the human which was reduced to the machine.

Perhaps a helpful exercise would be to return to Aristotle. In the Nicomachean ethics, he attempted to define what makes the human being distinctive against other animals and forms of life. Since then, we’ve learned a lot about the human, constructed immensely powerful technical systems, and the role of the human has changed immensely. Perhaps we should aim once again for what is distinctively human, not what is easily replicable or reducible. Let’s do things that are deeply related to human pursuits and those distinctive aspects of mind. Embrace the embodied nature of our experiences and skills, embrace the distinct cultural context that allows incredible displays of art. Write and think critically; push at the margins of ideas. Do it better than anything else can because of the peculiarities of where we are situated.

Such a conception of the relation between human and machine must involve the rich histories and culture which machines and electronically mediated forms of interaction have become an important part, but not the only one. It is through these histories which machines do not embody in the same fashion that we can return to writing and working not in a way that simply makes us intelligible to a machine, but in a way that embodies the richness beyond such means of comprehension. The machines are unavoidable, as the Luddites have discovered. That does not mean that they are or ever can be exhaustive.
Selected Bibliography


