Machine and Mind: Questions for Harry Collins

Walter B. Gulick

In a series of articles and books, Harry Collins has investigated the limits of what computers and robots can do and know. This is an issue of vital importance in the world today, and Collins is to be thanked for his exploratory forays into foggy territory. Through his efforts the fog has thinned. I appreciate and enthusiastically concur with his humanistic opposition to those who see no limits to the digitization and robotization revolution. But I think Collins' pioneering effort is too narrowly focused. To reach for greater comprehensiveness and clarity, I will raise questions for Harry Collins, questions that incidentally challenge the adequacy of some of the assumptions that guide his effort. For I believe some of his assumptions occlude key epistemic and ontological aspects of human existence from visibility in the comparison with robots. Then I will all too briefly offer suggestions regarding a more capacious understanding of the limits of artificial intelligence. I'll suggest the outcome gives credence to a revised threefold division of ways in which artificial intelligence cannot adequately emulate human intelligence.

I. <u>What exactly is Collins trying to determine and claim?</u> The answer to this question is not as obvious as might at first be supposed. Here is what Collins says in his 2010 work, *Tacit and Explicit Knowledge* (henceforth cited internally in the text as *TEK*): "The aim of the book is to reconstruct the idea of tacit knowledge from first principles so that the concept's disparate domains have common conceptual language" (*TEK* 2).[1]

Collins borrows the notion of "tacit knowledge" from Michael Polanyi, but it is important to see that he defines the term differently than Polanyi. Collins states that "'tacit' is knowledge that *cannot* be made explicit" (*TEK* 4). Polanyi recognizes unspecifiable powers inherent in all our acts of knowing, each of which integrates background information to form focal knowledge and intended action.[2] Such skills as bicycle riding, recognizing a face, or even the *process* of speaking coherently are typical examples of tacit knowing for Polanyi. From the standpoint of an experiencing knower, some aspects of the action of riding a bike (for instance, the focusing of eye muscles[3]) are tacit and not consciously experienced but some are likely explicit (for instance, seeing where one is going). From the perspective of the person undergoing an experience, all knowing has tacit aspects,

but from the standpoint of a scientist examining the experience, perhaps all aspects of the experience can be made biologically explicit (as Collins notes in his category of somatic tacit knowing).

Polanyi's notion of tacit knowing is consistent with the notion of doing things by second nature. It has some commonalities with Ryle's distinction between knowing how versus knowing that, or with Daniel Kahneman's intuitive thinking fast versus reason-based thinking slow, or even with Heidegger's present at hand. Perhaps most instructive in relation to the construction of robots, however, is the psychological distinction between *explicit* (or declarative) memory and *implicit* (or procedural) memory. Explicit memory, neuroscientist Eric Kandel states, "is the conscious recall of people, places, objects, facts, and events" while implicit memory "underlies habituation, sensitization, and classical conditions, as well as perceptual and motor skills such as riding a bicycle or serving a tennis ball."[4] To oversimplify a bit, when implicit memory is evoked and enacted, Polanyian tacit knowing occurs. The outcome of using language is explicit meaning, but the process of speaking or writing is a skill involving tacit knowledge. The fact that Polanyi regards all actions of knowing something as skill-based means that all knowing contains tacit elements.

A further characteristic of Polanyi's notion of the tacit is very important with respect to artificial intelligence. He states that those ideas or images that are the background information one uses in shaping focal thought have a tacit character that is different in kind from the experienced quality of what one focally perceives or thinks about. He calls this background material "subsidiary" to what one focally attends to.[5] When one makes explicit the subsidiary factors involved in knowing, you change these factors in a way that disrupts their character as known or performed. Thus when a pianist or typist starts focusing on finger placement, the holistic flow of a performance breaks down. In learning to play the piano or to type, explicit linguistic instruction may play an essential role in that learning process, but mature embodied knowledge, involving muscle memory, has become second nature no longer requiring focal attention. Implicit memory has replaced explicit attention. The accomplished concert pianist can attend to the meaning she is trying to express without having to focus linguistically on how she is performing.

A machine like a computer, on the other hand, is constructed top-down to achieve certain ends by following explicit instructions. Even robots having a degree of autonomy rely on electronically mediated signaling that, despite the sophistication introduced by massive connectionist processing and neural nets,[6]

arises from and depends upon explicable digital signaling. Polanyian tacit knowing, on the other hand, is the product of a long bottom-up process of evolution in which more and more complex ways of knowing emerge, culminating in human language usage. Language usage builds upon tacit comprehension we share with other animals, and tacit understanding often serves as the inarticulate standard in terms of which we measure the success of our articulation in language. We know more than we can tell.

One of the assumptions in Collins's *Tacit and Explicit Knowing* is that if something can't be made explicit, it can't be digitized and support artificial intelligence. His only candidate for what can't be digitized is what he calls collective tacit knowing (henceforth CTK). CTK is language-based social knowledge relevant to a specific context. This knowledge is derived from the process of socialization into some cultural domain. He claims the meanings of language and the sense of what is appropriate in different situations are too protean and unpredictable to be adequately explicated and digitized.

The above considerations lead to this paper's first crucial point. *Polanyi's understanding of tacit elements should constitute an even stronger category than CTK of what can't be digitized, because the very process of making subsidiaries explicit changes them so they are no longer what they were in the process of knowing.* The distinct levels of attention involved in expert piano playing—tacit muscle memory and explicit attention to the music's meaning—can't be reduced to one-level explicit digitization.

It is now time to consider Collins's objective in the second book of his being examined in this paper. His aim as stated in his 2018 book, *Artifictional Intelligence*,[7] (henceforth *AI*) is somewhat different than his attention to tacit knowing in his earlier book. He says, "The central theme of this book, then, is the embedding, or non-embedding, of computers into social context, and the difficult question is whether deep learning, in spite of its current and still greater potential ability to learn from the huge body of everyday human interaction, can do it; I will argue that in spite of the almost unbelievable recent successes, we are not there yet." (*AI* 5) The more recent book, then, is primarily about CTK[8] and is dependent upon what he claims to have achieved in his earlier book.

An objective somewhat hidden behind Collins's explicit statements of intention is to examine the likelihood of artificial intelligence progressing to the point where it is equivalent to or superior to human intelligence. This aim becomes most evident in Collins's inquiry into whether computers can progress from what

he calls Level II of artificial intelligence, "Asymmetrical prostheses," to Level III of artificial intelligence, "Symmetrical culture-consumers" (see *AI* 76-80). At Level II, we use the information and devices of artificial intelligence; they don't use us. At Level III, reciprocity between humans and AI devices may occur. But Collins claims such apparent equality can only occur if computers can be embedded in social contexts. With current electronic technology, he argues computers cannot be so embedded even using deep learning (*AI* 2).

Collins presupposes that the way to tell if machines can assume Level III status is to test them with new, more complex versions of the Turing Test. A computer is said to have effectively internalized knowledge if it is impossible to distinguish which of alternative answers to a series of questions are a computer's answers and which are human answers. An advanced Turing Test Collins proposes to check whether computers reach Level III is to see if it can repair broken language (*AI* 192—see *TEK* 115 and *AI* 3-4, 36, and 80 for good examples of what Collins means by a broken text).

Notice that for Collins the way to ascertain whether a computer has human-like intelligence is to determine whether its language-based knowledge is savvy in different contexts. He of course recognizes that skills like bicycle riding are examples of embodied knowledge, but in *TEK* he claims that skills are examples of "somatic tacit knowledge" (henceforth STK) which are essentially to be understood as mechanical in nature. We typically learn to ride a bicycle without knowing the physical rules that enable us to stay upright and make progress. But those rules can be explicated and employed in a robot.[9] When carrying out acts of second nature, like riding a bicycle or driving a car, "the human, per individual body and brain—that is, the human per complicated animal (that is, the human engaged in mimeomorphic actions)—is continuous with the animal and physical world. We are just like complicated cats, dogs, trees, and sieves." (*TEK* 104)

I believe three problematic beliefs are incorporated in this way of construing things. But first let me acknowledge a partial truth. Yes, in carrying out habitual actions that do not require conscious deliberation, humans act in ways consistent with many of the actions of other animals. But these actions employed by bodies are functional, achieving some purpose of a living entity, unlike how a sieve is inert unless employed by some agent. There is a difference between animal behavior, even when responding habitually to a stimulus, and mechanical events that only conform to the explicit laws of physics and chemistry.

Well, that gets us to the first problematic view of Collins. He sees the world in dualistic terms. On the one hand, humans are seen as rational, language-using beings capable of freely intending certain actions. But on the other hand, humans are also seen simultaneously to be like other animals, embodied beings driven by complicated mechanistic principles and digitizable laws of cause and effect. Collins thereby slips into the old Cartesian problem of understanding how thinking things are related to extended things. This leads to his second problematic view: that non-human animals are nothing but "complicated sets of mechanisms" (*TEK* 105). While science has not yet uncovered all the intertwined patterns of cause and effect in us and other animals, in principle for Collins they can be explained scientifically. Hence somatic tacit knowing can be fully explicated and captured by computers. But is a mechanistic view compatible with observed animal behavior?

The third problematic assumption involves the nature of knowledge. It leads to the second basic question I address to Collins.

II. <u>Is knowledge best regarded as an objective phenomenon?</u> Collins assumes knowledge exists objectively in the world in some sort of Durkheimian space, and that humans can capture some of it as explicit knowledge, but that some of it exists as tacit knowledge that can't be captured even in principle. There is much knowledge individuals don't know but which could be explicated if appropriate effort or explanation was made. Collins calls this unexplicated potential knowledge "relational tacit knowledge" (henceforth RTK—see *TEK* 86).

But wait a minute. Is knowledge person A has but conceals from person B or doesn't explain it thoroughly a matter of tacit *knowledge* for B? Isn't it simply a matter of B's ignorance? If person A knows something that B doesn't, this is not a matter of inexplicable tacit knowledge for B (given Collins's understanding of the tacit as that which is inexplicable); it just hasn't been made explicit to B. Of course information that some people know but others don't—the basic condition of human existence—could be computerized by those that know. This status hardly justifies the establishment of the category of RTK. Relational tacit knowledge is not even a weak form of tacit knowledge; it is just a person's lack of knowing something and should be eliminated as a category.

I side with those who believe knowledge is situated in minds, not in some abstract space. This claim about the location of knowledge, of course, takes a position regarding long standing issues of disagreement in sociology, and challenges the position Collins has taken. His view contrasts with the convictions of Polanyi, Dreyfus and many others who situate knowledge in a body and its

epistemic processes. My view, which I find beautifully articulated by Polanyi, is that knowing is a personal action, a skill, if you will. There is an object that one's act of knowing addresses. That object can be some perceived object, some claim made by a person or made impersonally in a text, some goal one can accomplish, and so on. Thus, for instance, the instruction in a manual for how to fly a type of aircraft was doubtless written by a knowledgeable author. But in itself the resultant writing is just a set of letters having the potential to inform a reader about a specific subject. That is, it is a mediating device having the potential to transfer information from one mind to another (as Collins gets at when he speaks of "strings"). The manual is not in itself a piece of knowledge. Instead knowledge is a person's state of embodied satisfaction regarding truthfulness arising in a questing embodied *mind* as it engages and evaluates an object of interest, such as a manual.

So, then, I reject the notion that knowledge is objective, although I will have more to say about this when discussing collective tacit knowledge in more detail. What of the process of knowing? When seeking to ascertain whether artificial intelligence is human-like, Collins regards knowing as either an intellectualistic social phenomena or as scientific understanding of causal processes. In each case, knowledge is dependent on language. This is evident in his reliance on the Turing Test and his version of Searle's Chinese Room test for assessing the human-like quality of artificial intelligence. Each relies on the ability of a computer to answer questions (use language) as a human would. Tacit skills, as non-linguistic, seem to be left out of the picture.

Let us return to the first two of Collins's assumptions that I have claimed are problematic. His problematic understanding of animal intelligence needs to be pondered further by addressing a third basic question.

III. <u>Do any of the more complex non-human animals possess some kind of knowledge, or are they merely complexly evolved beings little different in principle than machines?</u> Collins's Cartesian view of animals is not consistent with the vast amount of literature that attests to the myriad ways that animals engage each other and their environments with the same sort of attitudes and emotions that are expressed in human behavior. Consider, for instance, the following two general statements. Frans de Waal notes that animal studies

> have dispelled the notion that natural animal communication is purely emotional. We now have a far better grasp of how communication is geared to an audience, provides information about the environment, and relies on interpretation by those receiving the signals. Even if the connection with

6

human language remains contentious, our appreciation of animal communication has greatly benefitted from this research.[10]

Michael Tomacello writes,

> Thus, great apes, as the closest living relatives of humans, already understand in human-like ways many aspects of their physical and social worlds, including the causal and intentional relations that structure those worlds. This means that many important aspects of human thinking derive not from humans' unique forms of sociality, culture, and language, but rather, from something like the individual problem-solving abilities of great apes in general.[11]

Animals evolve through a kind of bottom-up adjustment to environmental affordances. Circumstances even within a normally comfortable niche differ, so living beings need mechanisms of adjustment if they are to prosper. Innate and learned responsiveness to signals bearing on an animal's survival are necessary. Humans share with animals some types of nonlinguistic reflexes, such as those associated with fight or flight. Kandel notes that "emotions are part of a preverbal system of social communication that we share with other animals. In fact, even with our extra facility for language, we use emotion every day to communicate our desires to one another and to monitor our social environment."[12] Machines can be designed to respond to specific signals as well. One thing that distinguishes "more advanced" animals from machines is the emergence of conception that allows for planning, for the consciousness of things not present. Here I find Susanne Langer's postulation of three types of meaning helpful. The *signals* transmitted by the body's receptors inform an animal of what is present and may need response. All animals, including humans, dwell in a world of incessant signaling. *Symbols*, however, allow for consideration of that which may not be present. For Langer, they come in two types: presentational (or what I will call *imagistic*) symbols and *discursive* symbols (based on language) unfolding in time.[13] Many animals appear to make use of imagistic symbols—imagination—for strategizing to find food, or to escape being themselves hunted for food. The point relative to Collins is that animals have complex tacit abilities that, like the subsidiary-focus distinction discussed earlier, involve more than one epistemic level. Animals are endowed with minds and are not merely complex one level stimulus and response machines. Humans inherit many of these tacit capabilities, and any thorough comparison of artificial with human intelligence ought to test whether such tacit skills can be found in robots or computers.

Collins, of course, rejects the use of semiotics or psychological categories like state of mind, claiming such internal factors are irrelevant to what counts in assessing artificial intelligence, namely, what it can do or say, what behavior it manifests. He believes only observable factors that are amenable to scientific measurement offer trustworthy evidence. However, just as behaviorism has proven to be a limited psychological theory incapable of explaining many facets of human behavior, so evaluating robotic intelligence in terms of behavior alone tells little about how close to human intelligence the robot is. To Collins's credit, his emphasis on the role of language and collective tacit knowledge is an important enrichment of analysis beyond behaviorism. Note, however, that his two categories of mechanistic behavior and linguistic social knowledge lodge us in the restrictive dualism I've claimed is a problem in Collins's thought.

Here we arrive at this paper's second crucial argument. *If, as Collins correctly claims, collective tacit knowledge represents shifting social practices that individuals adjust to intuitively but that elude the objective identification needed for adequate computer representation, then all other facets of human experience that are similarly indeterminate ought to be granted similar categorical recognition.* What is indeterminate can't be accurately digitized. And Polanyi persuasively argues, against the ideal of pure objectivity, that indeterminacy pervades all acts of knowing to various degrees.[14]

With respect to Collins's three categories of tacit knowing, I see relational tacit knowing to be an incoherent category, as argued earlier. Somatic tacit knowing seems useful insofar as it describes the potential for scientific understanding of such biological features as physiology, morphology, neuroscience, genetics, immunology, and the like. Collective tacit knowledge, too, is a useful category, although not, I think, in terms of the way Collins describes it as an objective feature of society. One of Collins's examples of CTK is how close people can be and still feel comfortable. No doubt there are cultural contributors to individuals' sense of acceptable nearness (or inclination to hug, or spacing in driving), but in addition genetic influences, previous experiences of intimacy, and a person's mood of the moment help make CTK so indeterminate. Collective tacit knowing seems better interpreted as a matter of individual interpretation riddled with different sorts of indeterminacy rather than as an issue of knowledge.

If the individual is the primary locus of knowing and interpreting, then all the tacit and explicit abilities of humans should be taken into account in comparing human and artificial intelligence. An enhanced Turing test can help facilitate language-based intellectual comparison. But since putting tacit factors into

language erases their tacit status, some other test is needed to see if artificial intelligence has mastered tacit skills. I would suggest that an enacted Coping Challenge might be up to the task of providing information about how a robot responds in comparison with a human to different unusual or difficult personal situations. Such a test might provide useful information relative to Satinder Gill's concern about how helpfully computers and humans interface in touchy situations.

In place of RTK, STK, and CTK, I conclude by delineating three broad categories representing domains in which individuals seek to know, yet in which embedded indeterminacies render digitization difficult if not impossible.

1. Collective Cultural Knowing (CCK) This category acknowledges Collins's recognition that the plasticity of social experience makes up to date, adequate computer responsiveness difficult. But the suggestion that this indeterminacy arises from a protean objective knowledge is rejected and replaced by a number of factors, some of which I described earlier. Perhaps if a computer incorporated an algorithm that accepted and processed a database only of material from social media of restricted groups from the past year, this problem might be attenuated.

2. Embodied Experiential Knowledge (EEK) This category picks up on the importance figures like Polanyi and Dreyfus attribute to the body in its various nonlinguistic functions. It is not a scientific category that basically involves biological explanation; STK takes care of that approach. Rather it is experiential. It postulates that the experience of various feelings and emotions contribute in important ways to knowing, but that such felt presence eludes digitization. It notes that skills involve simultaneously at least two levels of attention, only one of which might be explicit. It includes reference to the indeterminacy of spontaneous ideas and imagery that support knowing as well as outfit fantasy and dreams.

3. Personal Performative Knowledge (PPK) This category acknowledges the importance of the indeterminate personal element in knowing that Polanyi claims is involved in all knowing. It recognizes explicit as well as tacit contributors to personhood. How a person orients his or her life is indeterminate but of vital importance. Rationality, commitment to certain valued outcomes, being responsible as well as claiming rights—all such phenomena related to personal identity belong in this category.

A final question addressed to Harry Collins might be whether he asserts as I do the difference in kind between artificial and human intelligence. His answer is surely "yes," and I celebrate his humanistic insight. But the reasons for our similar affirmations are quite different, as I indicate in this paper.

Endnotes

[1] Harry Collins, *Tacit and Explicit Knowledge* (Chicago: University of Chicago Press, 2010). Charles Lowney provides a generally appreciative extended review of this book in his "Ineffable, Tacit, Explicable and Explicit: Qualifying Knowledge in the Age of 'Intelligent' Machines" (*Tradition and Discovery* 38:1 [2011-2012]), 18-37

[2] Further elaboration on Polanyi's notion of tacit knowing and then an analysis of different treatments of the tacit by Harry Collins, Neil Gascoigne and Tim Thornton, and Stephen Turner is found in my "Relating Polanyi's Tacit Dimension to Social Epistemology: Three Recent Interpretations" (*Social Epistemology* 30:3 [2016]), 297-325.

[3] Michael Polanyi, *Knowing and Being: Essays by Michael Polanyi*, ed. Marjorie Grene (Chicago: University of Chicago Press, 1969), 126-127.

[4] Eric Kandel, *In Search of Memory: The Emergence of a New Science of Mind* (New York: W. W. Norton, 2007), 132.

[5] A number of the essays in *Knowing and Being* describe and give examples of tacit knowing. Polanyi describes the subsidiary-focal relation as presenting a "from-to" structure of consciousness. There are many factors including language in the "from" dimension in contrast to the relative simplicity of focal meaning, so in a number of articles beginning with "Polanyi's Theory of Meaning: Exposition, Elaboration, and Reconstruction" (*Polanyiana* 2:4 -- 3:1 [1992-1993], 7-42), I have argued for a "from-via-to" structure of consciousness in which the "via" represents the articulation provided by language in everyday human consciousness. See Walter Gulick, "Polanyian Biosemiotics and the From-Via-To Dimensions of Meaning" (*Tradition and Discovery* 39:1 [2012-2013], 18-33) for a more accessible version of this formulation.

[6] See Jean Bocharova, "The Emergence of Mind: Personal Knowledge and Connectionism" (*Tradition and Discovery* 41:3 [2014-2015], 20-31) for a helpful discussion of neural nets in relation to Polanyi's critique of the neural model.

[7] Harry Collins, *Artifictional Intelligence: Against Humanity's Surrender to Computers* (Cambridge, UK: Polity Press, 2018).

[8] An important sub-theme in *AI* is indicated in the book's subtitle: we must protect ourselves from the manipulation and control that the internet and social media make possible.

[9] See Mihály Héder and Daniel Paksi, "Autonomous Robots and Tacit Knowledge" (*Appraisal* 9:2 [2012], 8-14) for a thoughtful exploration of whether autonomous machines can approximate human intentional action. They introduce their reflections by discussing what might be learned from what a particular bicycle riding robot can and can't do.

[10] Frans de Waal, *Are We Smart Enough to Know How Smart Animals Are?* (New York: W. W. Norton, 2016), 109-110. Donald Griffin, Jane Goodall, and Marc Bekoff are just a few of the better known writers among a great number of evolutionary biologists, cognitive ethologists, and social neuroscientists who would strongly disagree with the stance towards animal intelligence Collins takes.

[11] Michael Tomasello, *A Natural History of Human Thinking* (Cambridge, MA: Harvard University Press, 2014), 2.

[12] Eric Kandel, *The Disordered Mind: What Unusual Brains Tell Us about Ourselves* (New York: Farrar, Straus, and Giroux, 2018), 57.

13 Susanne Langer, *Philosophy in a New Key: A Study in the Symbolism of Reason, Rite, and Art*, 3rd ed. (Cambridge, MA: Harvard University Press, 1957). See especially Chapters III and IV.

14 Michael Polanyi, *Knowing and Being*, 120.