



Luca Cambiaso, *Virgin and Child*, c. 1570

Art and Artificial Intelligence

by G. W. Smith, (c)2014, 2015

The field of artificial intelligence has endured some false starts. In particular – and in conjunction with the computer mainframe era of the 50s and 60s – lavishly funded programs by the Western defense establishment to obtain accurate translations of Soviet documents yielded ludicrous results. The further result was the so-called “AI winter” of the 70s and 80s, during which funding for any type of AI research was hard to come by.

I mention this only to demonstrate that the field of AI is no monolithic juggernaut. To the contrary, it is a human enterprise which, like all others, has its varied approaches, and its varied successes and failures – and which the educated layperson can follow with some interest; *but doubt not the evolutionary mandate to endow the computer with human-like intelligence.*

Hence continuing progress in the field, and two examples of which have emanated from the laboratories of IBM: “Deeper Blue,” which, in 1997, defeated reigning world chess champion Gary

Kasporov in a series of matches 3½ to 2½; and the more recent triumph of "Watson" in a staged version of a popular TV quiz show.

These, however, have involved aspects of intelligence heavily dependent, in the case of both man and machine, on brute force computation and/or recall: the ability, in the first instance, to evaluate thousands of potential board positions, and to recall the key portions of thousands of previously-played games – Deep Blue, at the time of its victory over Kasporov, was ranked as the 295th most powerful supercomputer in the world in the famous "Top 500" listing[1]; and, in the second, the ability to recall and correlate thousands of mostly useless facts. At this point in the ass-over-teakettles rush of humankind into a techno future, hardly anyone now doubts the competence of the computer in data-intensive situations; as such, however, they are relatively uninteresting in human terms.

The occasion of the current essay is the coming to prominence of a new, and far more elegant, technique, and one which is thought to mimic the functioning of our biological computers: *deep learning*[2]. Its name implies two strategies: first, the "stacking" of a single pattern recognition algorithm, each layer of which presents in turn to the layer above it an increasingly abstracted "representation" of the data which it has received; and second, a recognition by computer scientists – with their new-found humility – that a sure way to inculcate the computer with intelligence is through the tried-and-true method of *learning*, and this by exposure of the self-tuning algorithmic stack to explicit or implicit "training sets"[3].

This is where the visual arts come into the picture. Most AI research, as exemplified by IBM's "Watson", has been in the field of *natural language processing*. Deep learning, on the other hand, has enjoyed its earliest and most spectacular successes in an area heretofore considered one of the most challenging for AI: *image processing*. Any two-year old, for example, can tell the difference between a cat and a dog[4], but this has been traditionally a steep climb for the computer; and likewise, simple visual recognition tasks – the so-called "CAPTCHAS" (which, by the way, is an acronym for "Completely Automated Public Turing test to tell Computers and Humans Apart") – are in wide-spread use to deflect Internet "bots."

Hold on to your hats, therefore, to learn that deep learning systems have achieved superhuman or near-human performance in several image processing tasks – *and with relatively modest computing resources*[5]: recognizing hand-written digits; recognizing traffic signs; recognizing the subjects of a diverse set of full-color photographs; and detecting cell features in biopsy slices.[6][7] (The first of these, not incidentally, offers a peek into the workings of a "stacked" algorithm: the lowest layer will typically take upon itself the mere task of detecting the "edges," or outlines, of the handwritten strokes within the two-dimensional array of pixels; the layer above it perhaps the task of sorting these into lines, open loops, and closed loops; the next layer perhaps the task of sorting the figures into the categories of mostly linear, mostly closed loops, and so on; and the top layer perhaps the task of distinguishing between a poorly-written "5" and "6" to produce the final identification.)

The members of the deep learning community have thus far kept their collective nose to the grindstone – even to the extent of avoiding an identification with "artificial intelligence;" and indeed, the development of a low-cost, automated system for evaluating biopsies is no mean feat. The members of that community will perhaps therefore groan in unison to discover that I will now be bringing back into the picture the question of the larger "human" dimensions of their work – but there is an important reason for it.

On the one hand, an ability to recognize hand-written digits represents little more of human interest than the ability of the computer to win a chess match or a quiz show; again, however, this capacity has been achieved with relatively limited resources. Despite their professed agnosticism, *the adherents of deep learning must suspect that a massive "scaling" of their algorithmic stacks might well give rise to one of the core features of a "strong" as opposed to a "weak" AI: the ability of a computer system to assimilate vast quantities of data, not only by way*

of being able to reduce its environment to a manageable set of features, but also by way of being able to prioritize those features in respect to whatever goals it may have.

Nor can we discount the possibility that deep learning research will help bring into existence a *superintelligence*: at present, the largest supercomputing cluster has the power of some *three thousand* Deep Blue machines[8]; and just as these systems are often dedicated to the extended running of a massive model of the earth's climate, or galactic evolution, it is not out of the question that such a system could be dedicated to a deep learning "super stack", and provided with, as its training set, the entire textual and visual contents of the Internet.

These computer scientists, in short, face the prospect – not unlike that faced by the physicists crouching upon the sands of Alamogordo – of helping to unleash upon the world an unimaginably potent force[9].

It might seem prudent, therefore, that our first experiments in setting off such an intellectual chain reaction be carried out on a smaller scale, *and with the single goal of determining which aspects of said algorithms might incline the computer toward pursuits related to the aesthetic, as opposed to a pursuit of mere intellectual capacity – the former of which are now recognized by anthropologists as helping to mark the boundary between a brute nature and some higher plane of existence*[10]. And when the time comes to "scale up" such experiments, it might seem prudent, further, to confine them to said supercomputer installations, given that these are typically under academic control, and given further that each is typically housed at a single location.

Google and Facebook, however, are apparently ready to "cry havoc, and let slip the dogs of war": a recent series of news articles[11][12][13][14] document the fact that Google, in particular, has launched what has been described as the "Manhattan Project of AI" – to be carried out, however, not in some carefully demarcated sector of sparsely populated New Mexico, but rather within that company's world-wide network of servers, and with the goal of creating a wide-ranging intelligence whose reach will extend to pretty much every desktop on the planet.

There is for Google, of course, a huge economic incentive: a search engine which can understand one's anguished query, and bring one to the exact product or service which can address it, is worth terabucks; and hence Google's rapid-fire hiring of deep learning experts, and acquisition of deep learning start-ups. The company has also employed Raymond Kurzweil as its Director of Engineering, and he has been cited in one of these articles as follows:

Google will know the answer to your question before you have asked it, he says. It will have read every email you've ever written, every document, every idle thought you've ever tapped into a search-engine box. It will know you better than your intimate partner does. Better, perhaps, than even yourself.[13]

I must confess that, on the one hand, I am like the Isabella of *Wuthering Heights*, swooning under the demonic influence of Heathcliff: I have been a Google devotee since its earliest days, have hundreds of personal documents entrusted as email attachments to its servers, and have long recognized the possibility that it is Google which might give birth to a true "global brain;" but now that the rubber has begun to meet the road, and now that their reckless, if not to say adolescent, approach has become clear – I am alarmed.

Nor am I alone. One of Google's acquisitions, DeepMind, has apparently insisted upon the formation of an ethics board as a legal condition for the deal; and one of that company's founders, Shane Legg, appears thus in *The Daily Mail*:

"Eventually, I think human extinction will probably occur, and technology will likely play a part in this," DeepMind's Shane Legg said in a recent interview. Among all forms of technology that could wipe out the human species, he singled out artificial intelligence, or AI, as the "number 1 risk for this century." [14]

Mankind, as always, is its own worst enemy; but let us see if we artists of the *visual* might not be able to "part the clouds"!

Unfortunately, however, we will need to plunge even more deeply into our comparison between speech and vision if we wish to have a truly comprehensive picture of the situation; and at this juncture, we might just as well make explicit a point to which we have already alluded: it now seems fairly certain that both human speech and vision are implemented within the brain in stack-like fashion.

How this might work in respect to vision we have seen already in our breakdown of a corresponding computer vision stack; and in respect to speech, something like the following layers can be identified: a bottom acoustic processing layer, which we share to some extent with other vertebrates, and capable of picking out individual sound features from a continuous input stream and responding to primitive signals of distress and so on; a layer above this one, elaborated during the language acquisition phase of early childhood, and capable of assembling phonemic sound features into words; a third layer, also elaborated during language acquisition, and capable of assembling words into meaningful utterances such as directives, questions, and statements of fact; and a final layer, elaborated during a developmental phase which roughly corresponds to formal education, and responsible for assembling and correlating a comprehensive and definitive set of such utterances.

Returning now to our analysis of the current computing landscape, I think it is fairly well established that the large commercial entities such as Google and Facebook will be focusing their AI efforts on natural language processing as opposed to image processing; and in seeking to illustrate their vision, nearly everyone involved has immediate recourse to the famous "Turing test."

This test, as it is commonly understood, is the ability of a computer to understand and answer arbitrary questions with the same facility in a natural language, and with the same general knowledge, of a typical human; but Alan Turing was a far more subtle – and much besieged – thinker.

As presented in his famous 1950 essay, "Computing Machinery and Intelligence,"[15] *the Turing test in fact focuses on the ability of a computer to rival a man at pretending to be a woman*; i.e., at any given time, there are only two contestants behind the curtain (and who communicate with an interrogator via teletype): a man and a woman, or a computer and a woman; the goal of the interrogator, with his questions, being to determine which is the woman, and which not; and success in the test on the part of the computer being defined as a performance equal to that of an actual man in confounding said interrogator in a number of such trials.

Properly understood, therefore, the Turing test has a marvelous focus on the subtleties of the human psyche; and *given that sexuality is deeply intertwined with aesthetic judgement*, it therefore represents something very much like the ability of the computer to become sensitive to those same *human* discriminations which I have already mentioned.

In short, let us thank whatever gods there may be that this seminal theorist had a larger experience, as we might say, of the human condition; for here, combined in one gentle individual, was not only the computational mind which broke the WWII "Enigma" engine, but also a mind which could imagine this snippet of dialogue between interrogator and contestant – and which snippet exhibits as well the link between sexuality and aesthetics:

Interrogator: Will X please tell me the length of his or her hair?

Contestant: My hair is shingled, and the longest strands are about nine inches long.

At present, however, the commercial interests – i.e., Google and Facebook – exhibit no dedication to such a sensitivity, despite their debt to Turing; but if we are willing to continue our digression regarding language and vision, *we artists of the visual have an opportunity to help inject a truly human perspective.*

Inasmuch as human *vision* is the most advanced of our senses, with its binocular, full color apparatus, and inasmuch as the visual channel has a higher "bandwidth" than the audible, it might be supposed that the former has emerged as the quintessential "human" modality – but both science and the humanities have reached the opposite conclusion: in the parlance of the deep learning community, it is the collection of words and utterances generated by our natural language processing capability which has emerged as the definitive "representation" of human experience, and this certified by both biology – i.e., those parts of the human brain dedicated to language acquisition, and culture – i.e., the status of the "word" as the ultimate repository of human wisdom.[16]

We practitioners of the visual arts may protest, and point to analogs – the vision centers of the brain, and the universal cultural understanding of certain visual patterns; the fact remains, nonetheless, that *the pioneer figure of Western culture is reputed by tradition to have been devoid of sight*[17]; and I can attest, in my own case, to a humbling fellowship – in Louisville, Kentucky – with the brilliant and mirthful community surrounding the American Printing House for the Blind[18].

What are the reasons for this extraordinary anomaly – the triumph of a less capable over a more capable modality? One, in particular – most obvious in retrospect, and therefore lost in the big picture: early hominids had the means for both the perception *and* production of speech; an efficient means of visual *expression*, on the other hand, did not exist for humankind until the relatively quite recent invention of paper.

How, then, are we to regard the cave paintings of, say, Lascaux? Without question, we are dealing here with both the most striking and the most convincing evidence for the appearance of humans like ourselves – and we are dealing as well with an extraordinary foretaste of the visual expression which would pour forth once paper, and canvas – and the computer screen! – became available[19]; by the time of these paintings, however, scholarship would suggest that the methods of oral-formulaic composition were already known to our early bards as a means of holding sway about the campfires[20].

And how, also, are we to regard the much earlier failure of evolution to follow up on the promise of *integumentary graphics*, as represented, say, by the species *Bothus mancus*[21]? Let us not be surprised, therefore, if the extraterrestrials with whom we first make contact are relatively mute, yet with enlarged foreheads able to display graphs of the formulae of physics – and images of their grandchildren!

We, however, are human. We can wrinkle our foreheads, or make them smooth; but it is in *words* that we must typically pour out the details of our hopes and fears. Genius that Turing was, this circumstance is the foundation of his famous essay, and which point I will illustrate by reproducing another of his segments of imagined dialog – and which segment again demonstrates his appreciation of the *aesthetic* as an essential ingredient of *human* intelligence:

Interrogator: In the first line of your sonnet which reads, "Shall I compare thee to a summer's day," would not "a spring day" do as well or better?

Witness: It wouldn't scan.

Interrogator: How about "a winter's day." That would scan all right.

Witness: Yes, but nobody wants to be compared to a winter's day.

Turing makes his point quite well, though without making it explicit: natural language encompasses the essence of what it means to be human, and of human intelligence. This, in turn, implies that a computer system aspiring to such an intelligence, and capable also of interpreting the raw speech of its human practitioners, must possess the capabilities, if not the exact functioning, of the human language processing stack; and if this, in short, is the challenge, then the newly elect of the deep learning community must be salivating in anticipation of a commercially-funded assault upon it.

Suppose, however, that there are inherent impediments to their implementation of a computer-based natural language processing stack; and suppose, further, that their heretofore quite successful experiments with image processing might – if extended – be more fruitful in terms of breaking into the realm of the truly human . . . ?!

In regard to said impediments, there can be no doubt – as already demonstrated by "Watson" – that computers can become frighteningly proficient in dealing with natural language; but anyone who has been exposed to the banality of a high-school debating society will understand that such a proficiency might well remain at some remove from the emotional and aesthetic intelligence which Turing had in mind – and which aspects of intelligence (I repeat myself) ought not be dismissed if it is our goal to achieve a "friendly" AI.[22]

So the bar has been set quite high; and in this connection, there are two related aspects of deep learning stacks which I have not yet mentioned: first, as the stack is dynamically exposed to its training set, the upper layers of a typical implementation send signals to the lower layers as to the effectiveness of their discriminations, and so the layers in effect grow together into a single unit; and second – as a corollary of the first – deep learning stacks tend to become "black boxes", and with the further tendency of their workings to become somewhat mysterious even to the computer scientists who have coded them [2].

Imagine, therefore, the challenge of duplicating the full range of capabilities – discursive *and* affective – of the human natural language "black box"!

To begin with, its various layers (to which we have already had some introduction) are embedded within the n-trillion neuron human biological computer as opposed to a laboratory computer system – so there is zero possibility, for example, of employing the typical software analysis technique of inserting a "HALT" instruction within the code which we are trying to deconstruct.

Of those several layers, furthermore, there is only one – the topmost, education-mediated layer – whose inputs are fairly represented by our much-heralded access to the texts of the Internet.

"This is hardly a limitation," the true believer might reply, "for most assuredly the complete syntax and vocabulary of a given language – i.e., that which is imparted during early childhood language acquisition – could be easily reconstructed from the mass of available texts even without the availability of grammars and dictionaries."

No doubt; but what can not be reconstructed from these texts is the steady stream of love and encouragement with which a mother accompanies her language training[23] – and absent which our computer system will have little chance of hearing the music behind the words.

And speaking of music, the emotive cries of the animal kingdom are no more than a step removed from it. Their influence, moreover, is still present within the brain's lowest, acoustically-oriented processing layer – and with a corresponding difficulty of access for the laboratory-bound computer scientist.

Consider, for example, the crisis-averting particle "OK," which has mysteriously emerged as perhaps the most universally understood and deployed human utterance, and with more than two

and one half billion Google hits to its credit. There are several etymological precedents, including the "Oll Korrect" of the Netherlandish proof-readers, and the "okeh" particle of Cherokee [24] – but must we not suspect that it is the echo of an ancient primate vocalization?

The above is an example of the patient working backwards that will be required if we are to endow a talking computer with the full range of sensitivities we associate with human speech; but the larger point is that there will be no "singularity" as it is currently imagined, i.e., a relatively quick and triumphant melding of human and computer intelligences – and here I present a comic analogy:

Members of the genus *Corvus* – the crows – are born with quite an innate intelligence, and are further subject to the influence of an elaborate culture which includes an extensive series of localized vocalizations[25]. We humans, nonetheless, must be to them as gods; but what team of ornithologists and computer scientists is prepared to put together a grant proposal with the goal of establishing a deep and enduring level of vocal communication with this black-feathered tribe?

All of which is not to say that there will not come a moment in the very near future when we recognize that natural language processing has crossed a certain threshold – yet the very phrase implies a *beginning* as opposed to a *consummation*.

Meanwhile, deep learning experimentation with image processing continues to gallop ahead, focused as it is on a more inchoate – and therefore perhaps more accessible and revealing – human modality; and here let me rush to my conclusion: *what if we were to establish something like a Turning test in visual communications*[26], i.e., one which would establish the ability of the computer to achieve a certain visual sensitivity?

The experiment I have in mind is one of simple binary discrimination, and is as follows: let us expose our algorithmic stack, as its labeled training set, to two collections of line drawings of the human figure – one consisting of "old master" drawings, and the second by amateurs; and then let us see, with a variety of subsequent drawings of similar origin, if it is possible for the computer system to perform a correct sort into the "master" versus "amateur" buckets.

This, of course, will be a test not only of computer science, but also of the entire edifice of art history and criticism: is there some objective basis for the judgements which we make in the name of art? And as confident as we artists are of a positive outcome, there remains the final objection that this is a measure of technique only – the more fluid line, and the more robust modeling, of the master artist – and therefore devoid of a larger significance.

We must grant the first term of this objection – but not the second.

Yes, technique is supposedly a matter of pressure and bearing only. The art lover, nonetheless, will claim that Michelangelo's ability to create lines of such great sensitivity was inseparable from his having been a "great soul," i.e., a person overflowing with reverence for the cosmos and all of its creatures; *and given that a similar paradox will be involved in endowing the computer with some non-trivial degree of empathy*, could not an approximation of our "visual Turning test" represent that breach in the wall through which computer science will end up pouring the bulk of its forces?

* * * * *

A final note or two – or, more properly, a *coda fantastique*:

As has just been implied, the problem of how an inanimate computer might manifest something like warmth and compassion is a subset of the question as to how these qualities arise within the human mind itself – which, after all, is said to be nothing more than a biochemical computer; and this, in turn, is a subset of the question as to how *any* degree of order and meaning has been

able to emerge from the swarm of fundamental particles of which the primeval universe was composed.

Here we have perhaps *the* great philosophical/scientific dilemma of the age; and if there is another which might possibly stand beside it, then surely we have reference to the incomprehensible *scale* of that universe – the rank upon rank of galaxies from the Hubble photos – in contrast to our own infinitesimally brief lives.

Yet Cambiaso's Virgin cradles her child with an untrammelled joy; and the child, in turn, holds out to her its tiny arms . . .

In attempting here, at the last, to tie together our various themes, this reference to the old master drawing by Cambiaso has an evident *initial* intention – to remind us, in a general way, of the key role that art must play in any attempt to approximate human behavior; but in noting that this is a work of art for which the word *poignant* might have been invented, some new and rather fertile connections become apparent: we are reminded first that, although art often deals with the great hero or the great event, another of its glories is its ability to elevate the quiet, the forgotten, the obscure – i.e., that with which our vaunted *compassion* must also concern itself; and at the other end of the spectrum, in discovering that "poignant" is from the Latin *pungere*, meaning to prick, or pierce, we will have suggested to us both the expanding bubble of the cosmos – and the importance of the smallest thing within it.

- [1] "Deep Blue (chess computer)," Wikipedia. (The reader will observe that I have herein made frequent use of Wikipedia as a source. I have been encouraged to do so not only by its wealth of material on the subject of AI, but also by my personal experience, as a novice contributor to the encyclopedia, of having encountered more than one dedicated, thoughtful, and patient computer scientist among its senior editors.)
- [2] "Deep Learning," Wikipedia.
- [3] The formal equivalent for "explicit or implicit" is "labeled or unlabeled", i.e., training sets in which members of the possible range of classifications are pre-identified as opposed to training sets for which a possible range of classifications is allowed to emerge spontaneously as a function of the particular set of algorithms employed.
- [4] This striking example is not original, but I have been unable to re-discover its source.
- [5] Given that today's cell phones have more computing power than the mainframes of a previous generation, the phrase "relatively modest" as it is used here might need some qualification. The fact of the matter is that the computing resources thrown at the typical deep learning trial could be considered obscene by historical standards – but in today's computing environment they are considered quite manageable in respect to the results being achieved.
- [6] "Jürgen Schmidhuber's Home Page," IDSIA [<http://www.idsia.ch/~juergen/>]
- [7] He, Kaiming et al., "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," February 6, 2015. [<http://arxiv.org/pdf/1502.01852v1.pdf>]
- [8] "Tianhe-2," Top500 Supercomputer Sites [<http://www.top500.org/system/177999>]
- [9] Urban, Tim, "The AI Revolution: Our Immortality or Extinction," Wait But Why, 2015. [<http://waitbutwhy.com/2015/01/artificial-intelligence-revolution-2.html>]
- [10] "Behavioral Modernity," Wikipedia.

- [11] Cadwalladr, Carole, "Are the robots about to rise? Google's new director of engineering thinks so...," The Guardian, February 22, 2014. [<http://www.theguardian.com/technology/2014/feb/22/robots-google-ray-kurzweil-terminator-singularity-artificial-intelligence>]
- [12] Gannes, Liz and James Temple, "More on DeepMind: AI Startup to Work Directly With Google's Search Team," RE/CODE, January 27, 2014. [<http://recode.net/2014/01/27/more-on-deepmind-ai-startup-to-work-directly-with-googles-search-team/>]
- [13] "Google will 'know you better than your intimate partner'," RT, February 23, 2014. [<http://rt.com/news/google-kurzweil-robots-transhumanism-312/>]
- [14] Prigg, Mark, "Google sets up artificial intelligence ethics board to curb the rise of the robots," The Daily Mail, January 29, 2014. [<http://www.dailymail.co.uk/sciencetech/article-2548355/Google-sets-artificial-intelligence-ethics-board-curb-rise-robots.html>]
- [15] Turing, Alan, "Computing Machinery and Intelligence," Mind, vol. 59, no. 236 (1950), pp. 4-30.
- [16] Speech, of course, can be reduced to the visual, as through reading, writing, and printing; but let us accept the premise of this essay, i.e., that natural language is essentially an *aural* phenomenon: language acquisition occurs well before reading enters the picture, and reading itself – as exemplified by our reading something aloud to ourselves to gain its full impact – can be thought of as a process of feeding pre-decoded words into the upper layers of the speech processing stack.
- [17] "Homer," Wikipedia.
- [18] Smith, G. W., *Aesthetic Wilderness: A Brief Personal History of the Meeting Between Art and the Machine*, 1844-2005, New Orleans: Birds-of-the-Air Press, 2011, pp. 42-43.
- [19] Nechvatal, Joseph, *Immersion Into Noise*, Ann Arbor: Open Humanities Press, 2011, pp. 72-89.
- [20] "Oral-formulaic composition," Wikipedia.
- [21] "Defensive strategy, flatfish," Youtube. I have always found this video somewhat disturbing. [<http://www.youtube.com/watch?v=teTbjE7VvhE>]
- [22] "Friendly artificial intelligence," Wikipedia.
- [23] "A Talk with Mom." [<https://www.youtube.com/watch?v=FR3y4TNGIjo>]
- [24] "Okay," Wikipedia.
- [25] "Crow," Wikipedia.
- [26] I am certain that there have been other proposals for a "visual Turing test," and my excuse for not tracking them down and citing them herein is quite simply exhaustion in terms of both the energy and column-inches which have been available to me in respect to this article; but should my own ideas gain some traction, there will be ample future – and much welcomed – opportunity for various synoptic approaches to the subject.