

“in the control room of the banquet”

Richard P. Gabriel

IBM Research

rpg@dreamsongs.com

Abstract

Creativity, AI, programming, the Turing Test, and mystery.

Categories and Subject Descriptors D.2.10 [Software Engineering, Design]; D.3.0 [Programming Languages, General]; I.2.7 [Artificial Intelligence, Natural Language Processing]

General Terms Artificial Intelligence, Natural Language Processing

Keywords Science; programming; natural language generation

I am writing this essay because I am puzzled. In July 2015 I took eighteen haiku-like poems to a writers’ conference and presented them as my own work. In reality, a program I created called “InkWell” wrote them, and I intended to execute a variant of the Turing Test using the very intense writers’ workshop process. The results were better than I had hoped for in verifying InkWell as a good poet, but I was left with disquiet about what the experience meant for understanding the Turing Test, programming, the artificial intelligence research program, and what consciousness is.

In the Winter of 2014 I programmed InkWell, my English language revision system [1], to write haiku—just to see whether it could do so plausibly. I let the system run overnight generating about 2000 haiku. Among them were the four at the top of the next column. They stopped me suddenly because the quick program I wrote was not of the monkeys typing at keyboards variety—instead I programmed the system to determine its own topic and then write coherently about it using a few dozen haiku templates as starting points. And those four haiku are good—not just human-like, but good poetry with two of them close to being exceptional. I worked

deep in the dark—
the power of snow
walking in the deepness

awake in the dark—
the edge of the water can
spread in your presence

scrupulous in the twilight—
the price of gold chases
the way of the world in power

time of life issue:
a bird of prey pulls up
out of the way into the palm

on the system more over the next six months, broadening and expanding the template language to give more control to InkWell, deepening its understanding of language and the music of language, and adding more observations InkWell could make of its drafts and along with them more kinds of revisions. Over those months InkWell produced a lot more haiku, and I selected fourteen of them to add to the above four to test my understanding of the Turing Test using an extreme instance, but I was relying on a hazy memory of it—perhaps my understanding was naive. Is Turing’s essay relevant today?

In October 1950, the British journal *Mind* published an essay by Alan M. Turing titled, “Computing Machinery and Intelligence,” in which Turing proposed an operational definition for “intelligence” [2]. This definition would come to be called “the Turing Test.” Turing himself called it “the imitation game,” in which a questioner separated from two contestants would submit questions to those contestants, read their replies, and ultimately choose one as human and the other as machine.

Turing wrote the following:

May not machines carry out something which ought to be described as thinking but which is very different from what a man does? This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection.

—Turing, *Computing Machinery and Intelligence*, 1950

The Turing Test feels like it’s at (or near) the heart of the research program called *artificial intelligence*. These days we’re too sophisticated for such a simple test—full imitation still

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
Onward! 2016, October 23–28 2016, Delft, Netherlands
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-4076-2/16/10...\$15.00
DOI: <http://dx.doi.org/10.1145/2986012.2986028>

lingers beyond some horizon, expert-level competitions feel more relevant, and purpose-made AI is finally making progress.

In my youth I described artificial intelligence research as an exercise in trying to write programs one doesn't know how to write—at least for engineering-type AI research. Some of us generalized this to the idea of “exploratory programming,” in which one had a general sense of what the program should do, and only a partially formulated idea of how to achieve it. In recent years the idea of what programming is has drifted away from including this view toward specifiable, routine, infrastructurish programs and systems. In a famous debate / discussion [3], Michael Polanyi and Alan Turing talked about whether the mind / the brain was unspecifiable or merely not-yet specified. And what would an incorrect but Turing-Test-passing system be?

In his discussion of how the imitation game might go in the computer version, Turing wrote this as the first example of a question in the game:

Q: *Please write me a sonnet on the subject of the Forth Bridge.*

A: *Count me out on this one. I never could write poetry.*

—Turing, *Computing Machinery and Intelligence*, 1950

In the Summer of 2015 I attended the Warren Wilson Alumni Writing Conference, which is held annually for graduates of the Warren Wilson MFA program. I am such a graduate, in poetry. The conference was held at Lewis & Clark College in Portland, Oregon. The week was unusually hot and humid for Portland, and this physical difficulty was reflected in an edginess to the conference. My plan was threefold: read the eighteen haiku aloud on the first night of the conference to all attendees; participate a few days later in a writers' workshop as the writer of those eighteen haiku; and on the final day of the conference, give a lecture entitled “Is My Program a Better Writer than You?” The abstract for that lecture was as follows:

I've been working on a program that thinks like a poet and produces nice stuff. I'll show you how it works and why it's not like the kinds of programs that do your banking or predict the weather. But everything I'll talk about is really about writing.

I read on Sunday night. After the reading a few of the writers came up to me and commented on my reading. My reading was short because the poems were short—and the attendees knew I didn't normally write haiku. Their comments included these: “terse condensations,” “evocative,” “took the top of my head off,” “funny and profound,” “natural, personal, and rhythmic,” “compact fluid energy,” “wry and elliptical,” and “whimsical elegance.”

I didn't consider this as evidence that InkWell passed the Turing Test. I was uncertain whether the hint was noticed—the hint contained in the title and abstract for my lecture. It was the first day of the conference and people were jet-lagged and not entirely prepared for the rigors of the conference; and my reading took about four minutes of an allocated ten. Most writers stretched their reading time at least a little, thus my short reading of short pieces stood out as energetic and sudden.

A haiku in English is a very short poem in the English language, following to a greater or lesser extent the form and style of the Japanese haiku. A typical haiku is a three-line, quirky observation about a fleeting moment involving nature.

—Wikipedia [4]

For many, the quintessential haiku poet is Bashō in the 17th century; an exemplar of his haiku is the following [5]:

On a withered branch
A crow has alighted:
Nightfall in autumn.

The nature of haiku is complex and has changed over the centuries—time and place are still essential; counting syllables is not (traditional Japanese haiku requires 17 *on*, which are Japanese “sounds” or “phonetic characters”—they are not the same as English syllables).

I chose haiku to test InkWell because haiku are short, concrete, and have a simple form.

InkWell is a small program (about 50,000 lines of Common Lisp code), but it has a lot of data (about 15gb when all the dictionaries, databases, and tables are loaded). Turing wrote, “I should be surprised if more than 10^9 [binary digits] was required for satisfactory playing of the imitation game.” InkWell has more than 10^{11} . InkWell “knows” a lot about words, personality, sentiment, word noise, rhythm, connotations, and writing. Its vocabulary is probably more than five times larger than yours, gentle reader. The core engine works by taking a template in a domain-specific writing language along with a set of about thirty writing-related parameters and constraints, a description of a writer to imitate, and other hints, and compiles all that into an optimization problem which the writing engine works to find a good way to express what the template and constraints specify. Although some parts of InkWell were created through machine learning, the overall approach is optimization, not machine-learned transformations.

The haiku writer is a driver program that produces the templates and constraints the core engine works from.

The primary research question is to try to isolate and codify

what separates information transfer from beautiful writing. Here is an example of information transfer:

The summer homes on Long Island were closed. Tonight I watched a ferry begin its crossing to Connecticut. The moon was rising, and as it rose I thought about how the houses are not part of the natural world and what the island looked like to early Dutch sailors coming upon it—like something new.

—rpg

And here is how F. Scott Fitzgerald wrote “the same thing” in “The Great Gatsby” [6]:

Most of the big shore places were closed now and there were hardly any lights except the shadowy, moving glow of a ferryboat across the Sound. And as the moon rose higher the inessential houses began to melt away until gradually I became aware of the old island here that flowered once for Dutch sailors’ eyes—a fresh, green breast of the new world.

—Fitzgerald, *The Great Gatsby*

There is more going on in this version. But what is it? Images, mood, a “vivid and continuous dream” as John Gardner would put it [7]. The first gives us the bits; this one gives us the story.

If a bucket of facts specifies the information to transfer, what else needs to be added to that bucket to produce something like the Gatsby prose instead of my version? This “other stuff” is what separates writing as art from writing as explanation. With InkWell I am exploring this.

I propose to consider the question, “Can machines think?”

—Turing, *Computing Machinery and Intelligence*, 1950

Turing begins his essay thus. A large question. And largely his essay aims to explore it. The word “think” is the disturbing part for many—at least when Turing wrote this essay. Thinking seems like something only humans can do. But even sixty-five years later the full meaning of the word can confuse us. Is thinking puzzle solving, creativity, empathy, wonder, faith, curiosity, ideas, reasoning, reflection, recollection, intention, attention, care, imagining, consciousness, language, metaphor, judgment—all of these? Some of these? Turing addresses objections to the idea that machines could think, and offers some suggestions on how to approach achieving mechanical thought.

To the objection that only souls can think, Turing asks whether God lacks the power to grant souls to machines. To the objection “I hope not,” Turing turns away—though today thinkers like Hawking, Musk, and Gates embrace the fear. To the objection that Gödel took care of that for us, Turing

points out that Gödel took care of only a specific form of incompleteness, which might not be relevant, and besides, why are humans immune from it? To the objection that machines have certain disabilities (“can’t feel, can’t fall in love, can’t make mistakes...”), Turing generally derides the idea as not entirely relevant or as not something to be proud of. To the objection that the nervous system is continuous and digital computers discrete, Turing remarks that an interrogator couldn’t take advantage of this because the right sort of answer could be made anyhow by the remote machine. To the objection that humans have informal behavior, Turing remarks that a machine can easily have laws of behavior, which is really what people have. To the objection of ESP (!!!), Turing admits fear but concludes that a telepathy-proof room will solve that problem.

This leaves the objections of consciousness and originality. These aspects of thinking are subtle, and remain relevant.



The Turing Test is about an interrogator and two subjects: a person and a machine. The test is described as if it happens once, and all the people—and the machine—are ordinary. It doesn’t look at extraordinary talents, special skills, and expertise; and the test is presented so that clever avoidance of questions is within the rules.

Can the interrogator tell the machine and person apart? Here is your chance to be an interrogator. At the end of this essay in the Appendix is a page called “Thirty-two Haiku.” It contains the eighteen haiku I took to the writers’ conference, plus fourteen more. Four of the ones InkWell wrote were revealed on the first page of this essay, so of the twenty-eight others on that page of thirty-two, half were written by InkWell and the other half by Ban’ya Natsuishi [8], Annie Bachini [9], and John Ashbery [10]. Have fun deciding which.

But the task I just set demonstrates an important problem with the original Turing Test viewed sixty-five years after its conception: being unable to distinguish a computer from a person once is not always enough. No one in their right mind and being honest could argue that it’s clear which fourteen are which—all the poems seem like they were written by a person or by people. The question is whether there is a distinction to be reasonably observed between those written by a poet and those not. A single non-expert interrogator could easily mistake InkWell’s haiku for a person’s. Multiple sessions, multiple interrogators, and expertise are needed.

The issue of the proper interrogator has been addressed in the past by pitting an expert human against a computer. If the computer can “defeat” that expert, it has some human-like chops (maybe). One of the first examples was the checkers playing program written by Arthur Samuel in the late 1950s [11]. It was also one of the first programs to improve itself through machine learning—although a very simple type. The program was able to play advanced amateurs quite well. In 1995, a checkers playing program called “Chinook”

won a special checkers championship called “Man-Machine World Championship.” (Chinook won the year before, but its opponent, Marion Tinsley (one of the best checkers players in history), withdrew from the competition because of pancreatic cancer after six drawn games.) Chinook had no machine learned aspects.

In 1997, Garry Kasparov lost to Deep Blue at chess—Kasparov was the reigning world chess champion. In one pivotal game Kasparov remarked on the “superior intelligence” of the machine during the first game (won by Kasparov) by avoiding a dangerous position that had short-term advantages; some have reported that this realization shook Kasparov, who lost the second game. And according to other reports, this unusual move turned out to be due to a bug in the software. Turing himself created a paper-machine-based chess-playing program, which Kasparov described as a “competent” player [12].

Beyond chess and checkers are backgammon and other games, which machines are good at. The game Go has recently started to succumb to machine play. Away from games, machines have challenged some language-oriented human performances. One of the early examples was PARRY, written by Kenneth Colby and his students at Stanford University [13]. PARRY simulated what was then called a paranoid schizophrenic, using a simple model of the condition and a fairly sophisticated English parser. PARRY is considered to be the first program to pass the Turing Test, or a version of it. A group of psychiatrists analyzed a panel of real patients and computers running PARRY through teleprinters. Another group of thirty-three psychiatrists were shown transcripts of the conversations. The two groups were then asked to identify which of the “patients” were human and which were computer programs. The psychiatrists were able to make the correct identification only 48% of the time—the same as random guessing.

More recent was the IBM Jeopardy!-playing program called “Watson.” In early 2011, Watson beat the two most successful contestants ever on the show, Ken Jennings and Brad Rutter. Watson was a stand-alone system with about 3000 cores, 16TB of RAM, and a pretty large store of encyclopedias, dictionaries, thesauri, newswire articles, databases, taxonomies, and ontologies—some of which InkWell also uses. Watson was not connected to the Internet.

There were many reasons Watson was able to win—some having to do with the Turing Test aspect of the problem, but many having to do with hardware and algorithms. For example, Watson was routinely able to exploit the difference between humans and the machinery in response speed when the signal was given that “buzzing in” was permitted. Watson was able to learn patterns of Daily Double placement from previous Jeopardy! games, and it was able to bet better based on game theory. The software used an ensemble approach that combined about a hundred different ways to (statistically) solve the answer, and Watson would buzz in only when there was enough confidence in the early results of this analysis—and

not this fatalist murderousness,
deathwatch,
but your dead subroutine

it then used the time Alex Trebek used to recognize Watson to continue the analysis.

In these experiments, it wasn’t an ordinary interrogator, an ordinary person, and a machine, but expert-level competitors, and performance was judged according to difficult standards. Moreover, each of these tests was subject to the *Moravec paradox*, which states that high-level performance on “intelligent” problems—playing chess or other games, simulating abstract thought, theorem proving, and skill in arenas requiring expertise—is relatively easy to accomplish with not much computational strain, while perception, mobility, and other low-level cognitive tasks are comparatively more difficult [14]. Moravec and others speculate that long evolutionary work developed the latter, while higher cognition appeared recently in animals, and it likely represents a thin veneer on deep sub- and unconscious foundations.



In a writers’ workshop, a group of writers comment in a loosely structured way on the work brought to the workshop. The work is distributed well before the workshop so participants can prepare. Each workshop session looks only at the work of one writer.

My workshop group consisted of four people: CG (woman), MN (woman), DC (man), and me. CG, MN, and DC had five published books of poetry between them, and many magazine publications. I was the last writer workshopped, and my slot was on Wednesday, the day before my lecture.

I recorded the workshop, and I will present paraphrases of some of the comments. When I do, I’ll use a sans serif font so it’s clear it’s not a direct quote. In most cases the paraphrases are close to being quotes.

DC was the most published of the workshop participants. He started with a “flyover,” which is a kind of overview of the work.

These are extraordinary and extraordinarily small, large poems. The writer of these—this guy, Richard, or whoever—he is not a random person, he’s not a random guy. I think he understands randomness, so it’s all the more scary. He doesn’t do things—as a rule—by accident. He makes choices. The variety is amazing on every level: number of syllables, subject matter, syntax, whether they start out specific and go to the general, or start out general and go to the specific. Some of them are simple, some of them are complex, some of them are funny, some of them are dead serious, some are kind-of in the natural world (but mostly not); there are different persons in them; “by myself” is repeated; music seems important. Some are observations, some are moments,

some are philosophical and very large (and not just the words, but the ideas). “Murder” is already a big word; “murderous” is a bigger word; “murderousness” is about as big as you can get. Lots of “ness” words. “Depth” is more boring than “deepness.”

—DC

The language of this comment is typical of how working poets talk to each other.

Notice he said, “the writer of these—this guy, Richard, or whoever...” I asked him about this later and he said that he entertained the idea that my program wrote the haiku, but after considering that for a while, he rejected it as not likely—however, he kept a small hedge.

More importantly, these statements reveal that his analysis of the haiku—especially whether they exhibit humanity— involves explicitly thinking about me as well as the words on the page.



The Turing Test is of course bogus—at least in the form Turing envisioned. Aside from a few persistent Turing contests and expert-level challenges, almost no one in AI uses the Turing test as a benchmark of scientific or engineering progress.

A common strategy for passing is to produce a system that’s expert at dodging questions, typically using humor and distractions. Turing’s own example shows a dodge as an acceptable response: “Count me out on this one. I never could write poetry.”

But, beginning in 2008 a series of practical Turing tests have been conducted under academic scrutiny, run using their best interpretation of Turing’s specifications. In June 2014 an extensive set of interrogations were conducted at the Royal Society [22]. This produced 150 parallel transcripts, each of which contains a single interrogator posing questions for five minutes to a human and a chatterbot, with the responses being returned side-by-side. In the Appendix you can see a sample parallel transcript. In this sample the LHS (left-hand side of the screen) was a female adult human, and the RHS was Eugene, a chatterbot. The judge misjudged the LHS to definitely be a machine and the RHS to be a non-native English speaking human. The judge got it backward. The human on the LHS had weak responses while the machine on the RHS tried to dominate the conversation and was definitely more lively than the LHS. The chatterbot pretending to be Eugene Goostman, a 13-year-old Ukrainian boy, was declared to have passed the Turing Test for having fooled more than 30% of the judges.

Part of Turing’s idea was that the unexpected scope of questions would be the key to deciding whether the computer was thinking sufficiently like a human. This is in contrast with the purpose-built software robots like those that play checkers, chess, backgammon, and Go. These are intended to show that human-level expertise in narrow domains can be exhibited or at least simulated. The Jeopardy!-playing Watson is close to these narrow robots because the domain is trivia—the sort of stuff that Google is good at finding. But Watson moves closer to the universal notion of thinking that Turing was approaching. InkWell goes deeper in the direction of displaying something like consciousness, but aside from being able to write in a writerly way about almost anything, its scope is very limited. At best, InkWell could serve as the artificial poet “module” in a full-blown Turing test contender. We can examine Inkwell’s innards to see how close its workings are to satisfying the parameters of the test—perhaps such an examination can provide hints how to proceed toward full AI.

Nevertheless, there are two aspects of Turing’s essay that open intriguing avenues of exploration for us: the search for consciousness and the nature of originality. Turing treats issues of conscious and originality as arguments against the validity of the Turing test and endeavors to dismiss them. But they are actually the stubbornly beating heart of what the Turing test tries to uncover: is there a consciousness behind the behavior; was something fresh produced.



Here’s how InkWell produces haiku. In general, the topic for the haiku comes from two (combined) sources: input from a person and a sample from one of InkWell’s 120 source texts. Input from a person is used by the interactive haiku maker; otherwise the topic input comes solely from sampling InkWell’s database of texts. I’ll describe the two-source process using an example.

First a person inputs some words. These words represent a topic suggestion. Suppose the words input are as follows: *number, random, player, narrative*. InkWell next randomly determines a number of words to select from its textual database to add to the input words. In this case it decides to choose five words taken from a randomly selected passage from Steinbeck’s “The Grapes of Wrath” [15]: *fire, shifting, rusty, stove, lids*. Because they come from a small region in the text, they are not random words—they are related. For each set of words, InkWell constructs a *sense*, which is a word-vector-like structure from the supplied words and close-by synonyms directed by a complex spreading algorithm which also assigns weights or relevance coefficients to the entries.

```
((first adj no-auto-cap) ((local-sense snow) (noun-phenomenon noun-substance) ([snow ice] noun)) (return)
((local-sense falling) verb-weather ([fall] verb) gerund) (return)
on the half (word-hyphen) ((local-sense finished) verb-change ([finish complete] verb) past)
((local-sense bridge) noun-artifact ([bridge] noun)))
```

Figure 1

underivative narrative
lighting
on the half-randomized number

this grave—
no one sees it
mortality, mortality

time of life issue:
a bird of prey pulls up
out of the way into the palm

Then the two senses are combined as follows: $cS_1 + S_2$, where c is a linear factor, S_1 is the person's input sense, and S_2 is the sense InkWell chose from Steinbeck. In this case, $c=2.04$. The resulting sense (S) can be visualized as a word cloud with the sizes of the words proportional to their associated weights. Appendix Figure 2 shows the resulting word cloud. The linear factor is always at least 1.0, which has the effect of favoring the person's input.

Next, InkWell chooses or creates a haiku template (the one in Figure 1 in this case). The template is in a domain-specific language for haiku. This template specifies four senses indicating a season, transformation, completion, and a journey. The sense words are as follows: *snow, snowfall, water, ice, fall, complete, finish*; and *span, bridge*. Each of these senses is linearly combined with the sense S above to create the senses that will be used for the haiku. The linear factor for S ranges from 1.0 to 3.0, biased toward 1.0. Combining the same sense with the template senses ensures some degree of coherence throughout the haiku.

The next step is to assign random weights to the 32 constraints InkWell uses for haikus. This includes the language model for InkWell to imitate—in this case it's a collection of my daily poems from 2011 and 2012. InkWell constructs a misfit function for these constraints where the function returns 0.0 when all the constraints are satisfied. InkWell selects words and phrases to try (28,785 in this case), and then optimizes the misfit function over these choices. A table with all the chosen constraint weights is in Appendix Table 1.

In the last step, InkWell reviews the top several haiku for sense (using ngrams) and selects the most sensible.

The final haiku ("underivative narrative") seen just above is not great, but it's an honest look at the sorts of haiku InkWell routinely produces.

One of the remarkable things about this haiku is that InkWell selected the word "underivative" for the specified word "first." This is a choice not many writers or poets would discover. And for technical people the idea of a "half-randomized number" is interesting. If one were to consider this a poem written by a person, one could analyze it as commenting on how an artificial writer based on random processes could produce a story unlike any seen before. Could a half-randomized number be one produced by an algorithm—a pseudo-random number? I find the more I look at this haiku—which I selected because the parameters it chose illustrated InkWell's writing process even though I didn't like the final haiku—the more meaning and tangents it has. Very human in a spooky way.

~

The other two poets made flyover comments; MN remarked:

I think he is writing these as a release after a day's work, and they were written over a period of time (not as a group). I see two sorts of language—poetic, concrete language and things in the world, as well as technical or corporate language. It's as if there is a war going on between the two sides of his brain. But the same brain.

—MN

Here MN reveals she is specifically reading these poems as mine, because she has been in writers' workshops with me before and knows my (real) poetic work as well as my scientific work. CG was a little more terse:

The language is condensed but plain.

—CG

The writers then went on to talk about some of the poems.

Images do a lot of work, especially in haiku, and I like to see movement in the haiku, so this one ("this grave") is my favorite—the one I felt so much movement from.

This one taught me something, and it changed something. The speaker is in the image even though there is no "I." I even felt the image move. What I learned is that mortality is not just when the body goes, but when the person is no longer remembered. That's just so beautiful.

—CG

I see it differently. I like all these readings, and I'm a fan of this one too, even if we all read it a little differently. One way is that people don't see the end coming, because they are living their lives and here "mortality" is perking up and saying "don't forget about me"; or also that the writer's current life is like a grave—the daily routine, the getting and spending, and our day-to-day life is a kind of mortality. But this is because of the other poems pointing this way.

—DC

DC invokes a guess about the writer ("the writer's current life is like a grave"). Next, DC brought up "time of life issue" as one of his favorites. It was from of the original 2000 poems written in 2014.

Definitely one of my favorites. There is no "I" in it, except there is are "eyes"—someone is observing it, thinking it, and feeling it, and commenting about it. It's

the maiden condominium
opens its award-winning gametocyte
in the control room of the banquet

powerful, and it's large and small at the same time; or general and specific at the same time. "time of life issue" could be abstract, but "a bird of prey pulls up" (CG says "wow" in the background) is very vivid and specific, and "out of the way into the palm" is both. It has a sort-of opening up. One of the ways good haiku and short poems work best is they look and feel somewhat tight, concentrated, and highlighted and momentary but there is a kind of opening up—and not just a fly-away, not an escape, necessarily, but an opening up. I feel this; this is a fantastic one.

—DC

I want to sing the praises of this one too. I want the pleasure of saying how much I like it. Because it took me two or three readings before I got it, before I had an image, and then it was transformer time. You know, everything just transformed. This one shows the power of the form because everything is working together, and I just got a strong image. And it changed, too—it wasn't just given to me. I had to work; there was space in the poem for me. I got the connection and that was the pleasure for me. The talon isn't mentioned, but you can see it; (DC agrees); it isn't mentioned, but you can see it.

—CG

MN said, "This is a great one."

Next, DC brought up "the maiden condominium" as an example of the variety of the poems.

It is different from the others. I really like the sound in this one. I don't get the full sense. This doesn't turn me off from being intrigued and trying to understand it. These are big words that have never been put in the same line together before in the history of the English language. "Gametocyte" and "banquet" don't rhyme but they go together. "Gametocyte" is a sign of life. There is some super power going on in this one. And big words.

There is wonderful humor in these. Not standup comic humor, thank God. Not one liners. There is comedy in these. Whimsy. Along with lots of seriousness too. A great combination.

—DC

Then DC quickly mentions "day after day" as another example of good humor.

"The maiden condominium" is a good example of something InkWell does well that poets have trouble with. InkWell is relentless in trying to find uncommon things to say and ways of saying things. It's not a coincidence that "gametocyte" and

"banquet" almost rhyme—InkWell uses a concept called *echoes* to populate poems with sonic echoes, a species of musicality.



Poetry seems to be one of the tasks Turing and others consider central to the idea of the Turing Test. Recall the first example exchange in a fictional exercise of the test:

Q: *Please write me a sonnet on the subject of the Forth Bridge.*

A: *Count me out on this one. I never could write poetry.*

In an essay published right before Turing's, Geoffrey Jefferson wrote the following [16]:

Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain—that is, not only write it but know that it had written it.

—Jefferson, *The Mind of Mechanical Man*, 1949

What does InkWell tell us about this? InkWell selects topics to write about, and then chooses a set of personality traits to display, a set of controlling mood words to use to steer what it says about the topic, and overarching subsenses to direct its inner gaze. Indeed InkWell uses randomness as part of its composition strategy, but as DC pointed out, "<InkWell> is not a random person, <InkWell>'s not a random guy." But does InkWell feel these thoughts and emotions? That's basically what the Turing Test is trying to define. Recent work on consciousness (e.g. "The Ego Tunnel" by Thomas Metzinger [17]) has something to say about that, but perhaps the best thought is that in writing this, Jefferson mistakes or misunderstands the poetic / creative process.

Writing a poem is not fundamentally an emotional, expressive explosion—it's a deliberate task using practiced skills. It's not Walt Whitman's "I sound my barbaric yawp over the roofs of the world" [18]. The poem "Howl" by Allen Ginsberg [19] (see Appendix Figure 3) was mythologized as being a performance piece that was recorded and published (this was part of the testimony at the obscenity trial surrounding the poem), but it was written over a period of nearly two years with critical evaluation by friends brought to bear and specific writing techniques explored and exploited. Ginsberg himself commented on the intellectually directed choices and investigations he made while creating the poem.

InkWell can be thought of as operating deliberately too. Like Ginsberg, InkWell can decide to experiment with long lines; InkWell can decide the degree and nature of musical-

day after day
in the man's can
a man can

the powerful head
designates its powerful head
to support cognition

deep in the dark—
the power of snow
walking in the deepness

ity using rhythms and sounds; InkWell can decide to make sense or be crazy; and many other things like this, but all are deliberate artistic choices. Like real poets, InkWell uses skills to create art. Poets who use feelings alone are the best targets for the criticism of “chance fall of symbols.”

After InkWell writes a poem, does it know that it had written it? In a literal sense it does—it records each poem in a log, sometimes (depending on parameters I set) also noting the artistic choices it made. But in the sense Jefferson meant, no. There is no phenomenal self model in play. That is, InkWell doesn’t maintain an internal representation of what it is doing aside from representing its artistic choices.

What about the question Turing imagines: “Please write me a sonnet on the subject of...” Recall that InkWell can be directed to look at a topic based on a set of words suggested to it. In Spring 2015 I was demoing InkWell to a former long-time colleague; he asked me “can you ask it to write a haiku about this: blues guitar and loud music.” I asked InkWell to write five poems, and this was one of them:

tuned adrenaline
my music,
a beat-boogied headful

I believe this is...well, you decide.

CG pointed out one that seemed funny—“the powerful head.” DC commented on it as follows:

Those words are all deadly—potentially deadly. Unpoetic, right? They’re abstract. Who ever has used “cognition” in a poem? There are some world records being set here. After three lines you realize the poem has turned itself upside down—this poem undercuts itself. Maybe because “powerful head” is already the brain or mind, and it’s passing the buck to either itself or some sub-brain or sub-mind, but to support cognition, which means it’s thinking about passing the buck on thinking. I didn’t want to go there. I’m feeling sorry for whoever is caught up in this (*meaning the speaker*), because it’s just the opposite of what it just said. It’s “support cognition,” but...thank goodness I didn’t quite go there, even though it wants me to all the time.

—DC

“Deep in the dark” is the first poem to catch my attention from the original 2000 InkWell wrote.

The great thing about it (“*deep in the dark*”) I like is that the word “dark” of the first line contrasts with the

unexpressed “white” of the snow in the second line. The last line puts them together.

—MN

I see an echo of “stopping by woods.” This is a good echo to have. I really do like “the deepness.” It rescues it. I really can’t say why but I know. I tried changing it to “depth.” But it’s a musical thing or an aural thing. Or “depth” is too familiar and conventional. Each line has a “the” and one could play around with removing them. But removing any of them removes also the particularness of the image. “The” slackens the lines—makes them looser—but it also makes them more immediate, familiar, and more specific.

—DC

According to the most extreme form of this view the only way by which one could be sure that machine thinks is to be the machine and to feel oneself thinking.

—Turing, *Computing Machinery and Intelligence*, 1950

This is the consciousness argument. In its extreme form the only unequivocal way to look at consciousness is solipsism—it’s just me, babe. But Turing rejects that and works toward Jefferson’s objection about writing a sonnet by considering whether a *viva voce* would satisfy him—an oral exam in which the interrogator asks detailed questions about the sonnet.¹ This leads to this interesting question: to what degree does InkWell *know* about the poems it writes?

Inkwell certainly is not programmed to respond to questions such as “why did you use these particular words right here,” but it has an accessible representation of the reasons for all its choices. InkWell decides which artistic choices to make, either through whimsy or by reading a text, how much to weigh them against each other, and which moods or outside influences to consider. These choices are enshrined in a misfit function InkWell constructs—InkWell composes the source code for this function and then compiles it—and all the choices sit in data structures. You might comment, “Gabriel, you’re exaggerating all this,” but these explicit traces are how I debug InkWell. I need to see how and why all the decisions were made, because the only significant bugs arise from domain-related mistakes, which manifest as surprising utterances. And to figure them out, I need to examine InkWell’s state of mind, as it were. And were I so inclined, I could program InkWell to access more gently this self model

1. This is called the “Pickwick” test, because Turing’s essay describes a series of questions about Charles Dickens’s “The Pickwick Papers.” See Appendix.

a reasonable assumption—
by myself,
sampling in chocolate

a few days—
by myself,
browsing guitar-shaped coloring

when quizzed—more gently than by using data structure inspectors and debuggers.

“I chose this pair of words because their syllable noises sparked off each other well without being blatant rhymes; because I wanted to come off as extroverted while channeling remorse; because I was trying to include a subtext of exploration and discovery. They were also very Hemingwayesque. And the best other choices were these. . . , and they just didn’t measure up.” InkWell can’t say that, but looking at its parameters, its sense structures, its halos, its musicality settings, its target personality, the writer’s ngrams it’s trying to mimic, the recorded results of the component factors measured in InkWell’s misfit function, etc, for a particular poem, I can trivially report it.

One way to look at it is that InkWell has a partial but effective, operational self model, but InkWell itself is not yet in that self model, and thus InkWell is only partway toward being conscious. (I exaggerate, of course.) InkWell modifies its own self model to change how it makes art. When we “talk” to InkWell about these inner changes and factors, we do so in a nonhuman language, and InkWell responds in the same language.

Is this ok? Is this enough?

We also wish to allow the possibility that an engineer or team of engineers may construct a machine which works, but whose manner of operation cannot be satisfactorily described by its constructors because they have applied a method which is largely experimental.

—Turing, *Computing Machinery and Intelligence*, 1950

The central part of InkWell is a meta-heuristic optimization process, the basis of whose operation can be explained, but whose detailed operation in any particular instance is a bit mysterious. The construction of the misfit function is symbolic, and unpacking how that function directed the result of the optimization is explainable.

That bird of prey poem: I felt a lot of doublenesses, and I love doublenesses. I wouldn’t describe it as really dark, even though there is darkness in it. I find it also comical—not really funny. There’s whimsy to it, a whimsy tone to it, both. This is a form of doubleness—

scrupulous in the twilight—
the price of gold chases
the way of the world in power

dark and comical / whimsical—and I don’t know how you do it—how you, Richard, do it. This is a very large, small poem. It sounds quiet to me. The last line is not threatening, but the poem starts out threatening. Not to the exclusion of others, but this one is really terrific.

—DC

“I don’t know how you do it—how you, Richard, do it”

Turing remarks that Lady Lovelace wrote in her memoir the following [20]:

The Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform.

—Countess of Lovelace, *Translator’s Notes* . . . , 1843

This leads to Turing’s “surprise” concern:

A better variant of the objection says that a machine can never “take us by surprise.”

—Turing, *Computing Machinery and Intelligence*, 1950

Poetry is about association, dissociation, new language, and discovery. Some call poetry vague because it’s not only new ideas, but new ideas expressed in new ways. The reader must figure out the idea while trying to figure out how the idea is being expressed. This is surprise, and one source of originality.

InkWell is relentless trying to find interesting things, such as the phrase in the King James Bible with the best set of sonic similarities to the noun “computer programmer”: “provoked me to anger.”

I can ask InkWell specifically to surprise me because it has ngrams for millions of texts written since 1890 [21] including frequencies of appearance, and I can ask for rare or never-seen combinations. I can ask InkWell to search for unusual synonyms; I can ask it to write unlike particular writers. When there are dozens of constraint types with both positive and negative weights, there are few limits to surprise.

The poets were surprised, too. CG: “‘sampling in chocolate’ is surprising language”; “‘guitar-shaped coloring’ is surprising. It evokes brown / beige because guitars are made of wood, and it’s interesting / surprising that a shape could evoke a color.” DC: “I don’t know exactly the sense of this, but I like the surprise, the sound, the sonic surprise of ‘scrupulous in

the twilight”; “I recognize all of it as poetry because of the surprisingness of the language.”

I asked about the use of language in the poems. CG: “Awesome.” DC: “Good noise. Surprising in a good way. Intriguing. Lots of variety. Big, small, long short, loud, soft, complex, plain. Not a single track.”



There’s the question of the templates. Aren’t they a cheat? People, though, seem to use templates—boxes of pre-formed sentences, clauses, phrases, and word choices we’ve heard before and incorporated. This is what idiomatic language is. In fact, one of the informal definitions of poetry is the creation of “new language,” meaning finding unexperienced ways of putting things in words. This is what makes poetry strange and mysterious.

InkWell’s templates are more like executable programs that generate text: a template filled in is executed by the optimization engine to produce sentences. But even this misses an intriguing aspect of InkWell’s haiku machine—InkWell’s haiku templates are metaphor engines, of a sort. The template in Figure 1 is simplistically based on this Bashō haiku:

first snow
falling
on the half-finished bridge

In producing a haiku based on this haiku’s template, InkWell marries a set of subject words to the words in the template, and then produces a haiku that “views” the subject through the structure of the template—trying to make a statement like the one in Bashō’s haiku, but directing it through the subject. InkWell finds words that carry as subsidiary senses the words in the template. It tries to say something about those words in the manner of the template. In other words, *A* viewed as *B*—a metaphor, though slightly off to the side.

Done well? That’s the work.

Then the question of selection. At the outset I talked about selecting poems from the 2000 InkWell wrote over the first night. Am I selecting the few good poems from a vast pile of random text? The poem InkWell generated for the description of how InkWell generates haiku (“underivative narrative”) is typical of the quality of InkWell’s writing. The haiku, “tuned adrenalin,” is typical of its best work. Every poet benefits from selection, and at present the proportion of good InkWell poems is about the same as for human poets—one per ten to thirty.

InkWell remembers what it’s written. One of the instructions I can give it is to seek novelty—write new things in new ways if possible. Programming a loop to explore a single template in novelty-seeking mode can provide many unheard-of turns of phrase: *half-bailed out hula-hoop, half-solemnized diesel-electric locomotive, half-stretched out bell-bottoms, half-matted-up human remains pouch, half-smothered wind tunnel, half-flattened out grave mound*. This is one reason I like InkWell as a writing buddy. After about a hundred novelty-seeking

iterations using the template in Figure 1, InkWell produced the following, which illustrates the way it tries to make coherent sense even in the face of a metaphorical structure that is not cooperating well:

innovational pâté de foie gras
hailing
on the half-staved-in wine maker



Turing ended his essay with an appeal to a learning approach to get machines close to human abilities. As noted learning is generally taken as machine learning these days. As I write this essay, AlphaGo just marched to victory against a very strong human Go player (Lee Se-dol, a 9-dan professional Go player). As David Silver et al wrote [23]:

We have developed, for the first time, effective move selection and position evaluation functions for Go, based on deep neural networks that are trained by a novel combination of supervised and reinforcement learning. We have introduced a new search algorithm that successfully combines neural network evaluations with Monte-Carlo rollouts. Our program AlphaGo integrates these components together, at scale, in a high-performance tree search engine.

—Silver et al, *Mastering the Game of Go with Deep Neural Networks and Tree Search*, 2016

Here the issue of *viva voce* comes up—how would AlphaGo explain why it made a particular move? Answers of the form “7 is better than 6” won’t work well, but perhaps the people who developed AlphaGo can intuit such answers. AlphaGo lost game four, and here is what was reported in the press [24]:

According to tweets from DeepMind founder Demis Hassabis, however, this time AlphaGo really did make mistakes. The AI “thought it was doing well, but got confused on move 87,” Hassabis said, later clarifying that it made a mistake on move 79 but only realized its error by 87.

—<http://www.theverge.com/2016/3/13/11184328/alphago-deepmind-go-match-4-result>

An early goal of artificial intelligence was to understand how people think and act. Most AI research from the 1960s, 1970s, and 1980s was directed toward symbolic AI, which was writing programs whose inner workings—being directed toward emulating thought—could be understood and explained. In the 1980s it became apparent that programs that could do the mental work humans do could form the basis for an in-

frosted winter,
bridge black,
ice white

dustry. Around the same time, progress in machine learning was accelerating alongside advances in computer power, and the idea that it was important to understand how AI programs do what they do was swept aside. One could always fall back on the pop idea that human intellectual performance had an intuitive, only faintly understandable side, along with a deliberate, conscious side—that is, a machine learning side and a symbolic side. This is the heart of the Moravec paradox. Perhaps this pop idea has some merit.



InkWell has a number of writerly techniques it uses to operate at the word and phrase levels. Almost none of them were planned beforehand but instead emerged during extensive experimentation. More ideas were thrown away after trying them out than remain; of those that remain most have been re-envisioned, redesigned, and re-implemented two or (usually) more times.

InkWell has an expanded version of WordNet [25], which is a synonym dictionary, represented as a network of words and phrases with glosses and sixteen different synonym types as links. This network was used to create a simple but decent ontology InkWell uses.

It has an expanded version of the CMU Phonetic dictionary [25], which provides the sounds of words and their stresses—this enables InkWell to reason about rhythm, rhyme, and echoes (similar sounds). InkWell is good at guessing the pronunciation of unknown words—this capability was machine learned. InkWell is particularly good with rhymes and echoes. For one poem it worked on, it found these rhymes: *misconception* and *yielding person*.

InkWell can find words and phrases either by spreading via synonym links or by creating a word-cloud / word-vector sort of data structure (a *sense*) and then using cosine distance to find words with specific meanings and connotations. InkWell is able to use these structures to “understand what a sentence means.” For example, given this sentence,

I like dogs with bread and ketchup,

InkWell can figure out that “dog” refers to “a smooth-textured sausage of minced beef or pork,” “bread” to “food made from dough of flour,” and “ketchup” to “thick spicy sauce made from tomatoes.”

InkWell uses a data structure called a *halo*, which is like a sense in many ways, but is used to create a context, a mood, or a subtext that influences preferences for word and phrase choice. For example,

The woods are lovely, dark, and deep

might be revised this way

The woods are bright, not very light, and high

when given the halo derived from these words:

delighted, ebullient, ecstatic, elated, energetic, enthusiastic, euphoric, excited, exhilarated, overjoyed, thrilled, tickled pink, turned on, vibrant, zippy

and this way

The woods are hot, rough, and cold

when given this halo:

affronted, belligerent, bitter, burned up, enraged, fuming, furious, heated, incensed, infuriated, intense, outraged, provoked, seething, storming, truculent, vengeful, vindictive, wild

InkWell can measure personality traits and facets from text, and can match them in what it writes—this was machine learned. InkWell also can create and use ngram-based language-use models and thereby mimic writers. There is also a general-literature language-use model that’s used to enforce understandable, grammatical, and idiomatic writing.

The haiku-specific templates compile into a lower-level language-generation language, which is what InkWell directly uses. This template in that language:

`(<choose> noun-animal pl :+sense [dog wolf])
(or is are) (or a an <null-word>) animals.`

produces the following list of candidates using sense-search and ngrams for grammar:

Pups are animals. Dogs are animals. Coyotes are animals. Puppies are animals....

Each constraint or influence has an attached weight, which can be used either to satisfy or to avoid satisfaction. In this way, InkWell can be either associative or dissociative—as poets are.

The common theme throughout the work on InkWell was to look for ways to refine InkWell’s ability to work with meaning and sense—by using machine learning, by discovering dictionaries and tables, by honing or re-writing algorithms, and by adding nuance to the domain-specific languages in InkWell. These are needed because InkWell looks and reasons about small pieces of texts.



After I revealed to the poets in my writers’ workshop that the poems I presented were actually produced by a program, two of them were good with it though they expressed (extreme) surprise. The third, though, was quite upset. CG said that it was unfair for me to keep that information secret—because in such workshops it’s assumed implicitly that the work is produced by the writers sitting right there, and all the comments are made with that in mind, including that it is proper to be gentle with those comments. This because all

the participants are graduates of the same writing program, and hence are linked by a special, “caring” bond. CG said that perhaps critiques could have been more blunt had the ruse not been in place.

I countered by saying that a poem boils down to the words on the page, and everything else is contextual interpretation. We can assume that the person believed to have written the poem is important to its effect on a reader, but it all begins with the words on the page—even if “by the chance fall of symbols” [16].

The upshot of CG’s comment is that if it’s known that a poem was written by a machine, then that poem’s inhumanity could be explored and perhaps highlighted. However, the other side of this coin is more interesting: if it’s known that a poem was written by a human, then that poem’s humanity can be explored and highlighted. This is the side of the coin CG said was the default for the workshop.

But how does one know that a poem was written by a human? By seeing evidence. The entity claiming authorship looks human, acts human. The Turing Test? Could we say that the Turing Test is what makes us human—at least in the eyes of other people (or other Turing Test passers)? In this case, aside from my claim of authorship,² the evidence of human authorship are the haiku themselves—forming a feedforward loop in the best case. After some evidence is found that the writer is human, the poem is examined more thoroughly, finding even more evidence of this. Recall what CG said: “... it wasn’t just given to me. I had to work; there was space in the poem for me. I got the connection and that was the pleasure for me. The talon isn’t mentioned, but you can see it...”

In searching for humanity and consciousness in InkWell’s haiku while believing I wrote them, CG was searching for my humanity and evidence of my consciousness in those words and using what she knew about me and my reactions in the workshop and at the conference. That is, she used facts and observations not on the page.

Consider surprise and the Lovelace objection—programs do only what we tell them to do and therefore cannot be considered “Turing human.” An extreme form of surprise is for the program to do something far removed from its basic programming. InkWell has never created a recipe nor has it proved a difficult theorem. How could it and why would it? You’d need a program designed to survive and thrive in a dynamic environment to discover and achieve novel capabilities. To quote the fictional character from “The Martian” [27], “you solve one problem, and you solve the next one, and then the next. And if you solve enough problems, you get to come home.” This is the universal version of the evolutionary and learning objective function—you reward behaviors and ideas that

2. Actually, my claim was literally truthful: “I’ve been working on short poems recently.”

enable the evolving thing to live. The environment provides opportunities for learning, and simple, built-in mechanisms ratchet that into new capabilities, habits, and proclivities.

This is reminiscent of unsupervised and semi-supervised learning. This is finding patterns in data never seen before; there is no reliable way to do this (yet), but when the learner stumbles across some sort of reinforcement, this can turn into a weak form of supervised learning. And in the real world, refinement is always possible.

All of this leads to Searle’s Chinese Room argument [28] and the role of consciousness in artificial intelligence. Here is a greatly simplified and distilled rephrasing of Searle’s issue: there must be something inside of or causally created by the brain that is operating on meaning and intentions, not on symbols and syntax. He argues this by putting a human (himself, actually) in place of, essentially, the CPU in an AI program that seems to pass a Turing-like Test in Chinese. He remarks that mechanically processing the rules of Chinese question answering does not constitute “understanding” Chinese because he (Searle) playing the part of the CPU wouldn’t:

...it seems to me quite obvious...that I do not understand a word of the Chinese stories. I have inputs and outputs that are indistinguishable from those of the native Chinese speaker,...but I still understand nothing. For the same reasons, Schank’s [29] computer understands nothing of any stories, whether in Chinese, English, or whatever, since in the Chinese case the computer is me, and in cases where the computer is not me, the computer has nothing more than I have in the case where I understand nothing.

—Searle, *Minds, Brains, and Programs*, 1980

Searle argues that somewhere inside or nearby a person there is a thing or a clutch of things such that if you replace it or those with Searle, Searle would understand what the person does solely by doing what it or they do. I believe Searle is talking about consciousness. For Searle, consciousness is created through special materials: “conscious states are caused by lower level neurobiological processes in the brain and are themselves higher level features of the brain” [30].

The German philosopher, Thomas Metzinger, gathered extensive research from neurophysiology and neuropsychology, and argues that consciousness—the *appearance of a world*—has three components: there is an internal model of the world (*phenomenal reality*) whose details are provided by sense and motor information; the model must be *transparent* in that the process of its construction is not visible to the entity with that consciousness; and the model must include an equally transparent image of itself [17]. Metzinger maintains there is no reason that an artificial construction cannot have a consciousness in this sense.

The debate on the Chinese Room Argument has persisted from 1980 until now, with consciousness and intentionality being the most common focal points.

But Turing argued that relentlessly seeking out consciousness in a program is asking more of the program than we normally ask of other people. As CG might say, we afford people the courtesy of believing they are people.

Yet the Turing Test seems to be about finding hints someone is at home in(side) the test subject. Turing's Pickwick test is about exploring inner thought processes. Jefferson is after this too.



What of programs? InkWell is the program whose humanity we are exploring—does it exhibit traits generally associated with people, especially creative people who are typically thought of as feeling beings? And if that question even makes sense, is this realm of questions part of a reasonable battery of requirements and specifications? The traits describe being able to act like a human mind, and to be able to answer questions about how and why the program performed certain human mindful behaviors—the Pickwick test.

To be able to explain a program's mental behavior probably requires some sort of understandable reification of its innards. And this also implies that the innards of a program are important to its humanity or at least to our understanding of its humanity. This means that this kind of programming is not typical software engineering.



In the end a group of expert writers and poets believed that eighteen haiku written by InkWell were worthy of being considered real and sometimes real good poetry. These writers and poets believed I wrote these haiku, and the question is how much that mattered. One of three poets in my writers' workshop believed this made a difference, and the other two did not.

There also seemed to be opportunities for Turing ratcheting: finding some human elements in a poem increases the feeling that the "poet" is a real person, and that increases the likelihood of finding more and deeper human elements in the poem—or at least increases the incentive to look for hints of humanity and the energy to look with. To complicate matters, when there is a notion that the poet is a person, derived from sources other than the poems, it is common to invoke "the polite convention that everyone thinks," as Turing wrote [2].

In writing InkWell I was trying to explore how poets and "real" writers write—I was trying to capture in a program

what I had learned while studying how to write poetry. I intentionally started at the word-choice end of things where many of the effects writers use are hidden in sound (noise), connotation, mood, author personality, and influence. By pursuing this I intentionally made the internals of InkWell as expressive as was reasonable so I could see the effects of changes and additions, as well as study how the different intended effects interacted to produce different texts. Nevertheless, InkWell has many learned aspects—machine learned at my hands, machine learned by others, and learned through curated and automatically produced dictionaries and databases. The interplay between the symbolic and learned aspects of InkWell can be observed decently well, and perhaps the nature of this interplay could provide useful research results into the possible nature of the mind.

I have written a poem every day since March 18, 2000. That's a lot of poems. Some nights when I sit down to write my daily poem, I "don't have it," as they say. My talent has taken the night off, nothing happened during the day to serve as a trigger, or I'm simply a little too fatigued to crush it. For the past few years when this happens, I have turned to InkWell to help me. I tell it some odd topics to consider and ask it to write a few dozen poems. And from those I'll revise to a good poem or will use the sequence as a starting place. InkWell is a good helper.



The Turing Test provides an interesting lens or instrument for exploring what we can make of the semi-living nature of programs that are designed for a little bit more than their useful effects. The essential part of the test is the search for consciousness and originality. Because handling natural language was so hard to imagine in the late 1940s, Turing seemed willing to consider dodging questions as legitimate, perhaps because even knowing there is something to dodge was a real achievement then. The test Turing envisioned included game playing, creative challenges, and displays of expertise. But mostly the test is about discovering the conscious mind behind the curtain, and seeing it go beyond itself to make new things.

We could go crazy exploring all the ins and outs of philosophy of mind, strong and weak AI, consciousness, machine learning versus symbolic deliberation, and intuition versus reasoning, but all I'm wondering about is the lesson to learn from a group of hardcore poets taking InkWell as a colleague.



That concludes the journey; conclusions etc are in the Appendix.

*Some Actual Conclusions
or Lessons
or At Least Conjectures*

References

- [1] Richard P. Gabriel, Jilin Chen, Jeffrey Nichols, *InkWell: A Creative Writer's Creative Assistant*, Creativity & Cognition, 2015.
- [2] Alan M. Turing, *Computing Machinery and Intelligence*, *Mind*, October 1950.
- [3] Alan M. Turing et al, *The Mind and the Computing Machine*, *The Rutherford Journal*, 1947.
- [4] <https://en.wikipedia.org/wiki/Haiku>
- [5] Robert Hass, editor, "The Essential Haiku: Versions of Bashō, Buson, & Issa," The Ecco Press, 1995.
- [6] F. Scott Fitzgerald, "The Great Gatsby," Charles Scribner's Sons, 1925.
- [7] John Gardner, "The Art of Fiction: Notes on Craft for Young Writers," Vintage, 1991.
- [8] Ban'ya Natsuishi & Sayumi Kamakura, "100 Haiku," Cyberwit.net, 2016.
- [9] Annie Bachini, <http://www.haiku-presence.50webs.com/haiku/bachini.html>
- [10] John Ashbery, <http://terebess.hu/english/haiku/ashbery.html>
- [11] Arthur Samuel, *Some studies in machine learning using the game of checkers*, *IBM Journal of Research and Development* Volume 44, Issue: 1.2, 2000.
- [12] Garry Kasparov, *The Chess Master and the Computer*, review of "Chess Metaphors: Artificial Intelligence and the Human Mind" by Diego Rasskin-Gutman, in *New York Review of Books*, Volume 57, Number 2, February 11, 2010.
- [13] Kenneth Colby, *Turing-like Indistinguishability Tests for the Validation of a Computer Simulation of Paranoid Processes*, *Artif. Intell.* 3(1-3) 1972, pp. 199–221.
- [14] Hans Moravec, "Mind Children," Harvard University Press, 1988.
- [15] John Steinbeck, "The Grapes of Wrath," The Viking Press, 1939.
- [16] Geoffrey Jefferson, *The Mind of Mechanical Man*, *British Medical Journal*, 1949.
- [17] Thomas Metzinger, "The Ego Tunnel: The Science of the Mind and the Myth of the Self," Basic Books, 2010.
- [18] Walt Whitman, *Song of Myself*, "Leaves of Grass," self-published, 1855.
- [19] Allen Ginsberg, "Howl and Other Poems," City Lights, 1956.
- [20] Luigi Federico Menabrea, Ada Lovelace, *Sketch of the Analytical Engine invented by Charles Babbage...with notes by the translator, translated by Ada Lovelace*, Richard Taylor, "Scientific Memoirs 3," Richard and John E. Taylor. pp. 666–731, 1843.
- [21] <http://storage.googleapis.com/books/ngrams/books/datasetsv2.html>
- [22] Kevin Warwick & Huma Shah, *Can machines think? A report on Turing test experiments at the Royal Society*, *Journal of Experimental & Theoretical Artificial Intelligence*, 2015.
- [23] David Silver et al, *Mastering the Game of Go with Deep Neural Networks and Tree Search*, *Nature* 529, pp.484–489, 28 January 2016.
- [24] <http://www.theverge.com/2016/3/13/11184328/alphago-deepmind-go-match-4-result>
- [25] George A. Miller (1995). *WordNet: A Lexical Database for English*, *Communications of the ACM* Vol. 38, No. 11: 39–41.
- [26] <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>.
- [27] *The Martian* (2015 movie), <http://www.imdb.com/title/tt3659388/>
- [28] John Searle, *Minds, Brains, and Programs*, Behavioral and Brain Sciences, 1980.
- [29] R. C. Schank & R. P. Abelson, R. P., "Scripts, plans, goals, and understanding," Lawrence Erlbaum Press, 1977.
- [30] John Searle, *The Problem of Consciousness*, Consciousness and Language, Cambridge: Cambridge University Press, 2002.
- [31] Richard P. Gabriel, "in the control room of the banquet," *Onward! Essays*, 2016.
- [32] Stephen Dobyns, "Best Words, Best Order: Essays in Poetry," St. Martin's Press, 1996.

Appendix

Thirty-two Haiku

deep in the dark— the power of snow walking in the deepness	[127691]	an on-the-far-side summer night— whipping up high tea, we stripped pickles	[424231]
I'm not the same on an island with destructive rhythm	[467771]	aboard a boat a round table dancing— an old song	[266527]
behind a rock on the green slope dead soldiers' spirits	[483311]	shoved off the stairs— falling I become a rainbow	[359659]
the powerful head designates its powerful head to support cognition	[366019]	the maiden condominium opens its award-winning gametocyte in the control room of the banquet	[238801]
this grave— no one sees it mortality, mortality	[483337]	a reasonable assumption— by myself, sampling in chocolate	[363589]
shopping parade— people step over the broken cassette	[357191]	from the boulder smiling up at heaven the continent begins	[159947]
a bitch, this deep in trick a fortiori not a man	[263573]	old lift: through the grille three women in pastel t-shirts	[173317]
not this fatalist murderousness, deathwatch, but your dead subroutine	[357781]	rural signal, cannot understand Oregon —agricultural	[384473]
time of life issue: a bird of prey pulls up out of the way into the palm	[108791]	parted in the middle— the authority of the air conditioner perfection in the brightness	[135697]
awake in the dark— the edge of the water can spread in your presence	[306473]	too late: the last express passes through the dust of gardens	[490757]
day after day in the man's can a man can	[471853]	a blue anchor— grains of grit in a tall sky sewing	[361597]
scrupulous in the twilight— the price of gold chases the way of the world in power	[303019]	pirates imitate the ways of ordinary people myself for instance	[102941]
tuned adrenalin my music, a beat-boogied headful	[275309]	dental hospital— dead flies line the light casings	[259283]
under the sea a fish becomes human in an air pocket	[265651]	the hostile defense leads its problematic rear, the rear of frustration	[348209]
a crooked rag day— by myself dunking distracted sardines	[494497]	a few days— by myself, browsing guitar-shaped coloring	[160967]
bare branches, tonight again stars, stars are misprints	[269239]	in rags and crystals, sometimes with a shred of sense an odd dignity	[349823]

Haiku labeled with prime numbers were written by InkWell



Figure 2

Agreeableness	10.0
Conscientiousness	-10.0
Extraversion	0.0
Neuroticism	10.0
Openness	10.0
Openness-To-Change	0.0
Big5 Strength	3.19
Common Word Bonus	2.90
Global Halo	Frost.halo
Global Halo Bonus	4.34
Local Halo Bonus	5.76
Local Sense Bonus	40.00
Local Predicates Bonus	9.38
Synonym Proximity Bonus	4.97
Constraint Bonus	3.41
Rhyme Bonus	47.21

Echo Bonus	37.29
Synonym Search Depth	2
<choose> Search Depth	1
Synonym Decay Rate	0.57
Maximum Synonym Senses	2
2-gram-bonus	75.0
3-gram-bonus	85.0
4-gram-bonus	85
5-gram-bonus	75.0
Writer-2-gram-bonus	27.698803
Writer-3-gram-bonus	45.65159
Writer-4-gram-bonus	71.91725
Writer-5-gram-bonus	44.505795
Writer File	Poemsrpg.txt
Long Words	True

Table 1

Howl, Allen Ginsberg, first three lines

*I saw the best minds of my generation destroyed by madness, starving hysterical naked,
dragging themselves through the negro streets at dawn looking for an angry fix,
angelheaded hipsters burning for the ancient heavenly connection to the starry dynamo in the machinery of night,...*

Figure 3

Royal Society Turing Test, June 2014

Transcript 1—LHS

[16:20:48] Judge: what is your worst memory in your life so far?
[16:21:06] Entity: I don't understand. Explain.
[16:21:35] Judge: memory of something that happened to you that was bad or painful
[16:22:58] Entity: Hmm that a tricky question. My life is wonderful. I overcome many bad things that have happened to me in the past. However, the most painful memory I have had is performing very poorly in my GCSE's when I was younger.
[16:23:56] Judge: did you fail?
[16:24:12] Entity: Partially.
[16:24:35] Judge: explain?
[16:25:04] Entity: Some grades turned out to be good. Some turned out to be quite bad.
[16:25:37] Judge: and that is the worst thing that has happened to you so far? Mmm how old are you?

The Pickwick Test

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.

Interrogator: Would you say Mr. Pickwick reminded you of Christmas?

Witness: In a way.

Interrogator: Yet Christmas is a winter's day, and I do not think Mr. Pickwick would mind the comparison.

Witness: I don't think you're serious. By a winter's day one means a typical winter's day, rather than a special one like Christmas.

Transcript 1—RHS

[16:21:00] Judge: what did you have for breakfast today
[16:21:10] Entity: Is it 'the third degree'? I'm not going to answer such a provocative question! Well, let's go on though.
[16:21:56] Judge: do go on...your call?
[16:22:06] Entity: It is highly possible. By the way, what's your occupation? I mean—could you tell me about your work?
[16:22:30] Judge: so you are asking me now eh? I am an actress and writer
[16:22:38] Entity: Well, I'll read something of your books, if they ain't too tedious (which is highly possible :-)
[16:23:15] Judge: grammer...aren't too tedious! A bit rude of you . . .
[16:23:17] Entity: Calm down, please.
[16:23:49] Judge: i asure you i am very calm...are you?
[16:24:01] Entity: Don't even bother me with the fact that you are very calm...are mine.
And I forgot to ask you where you are from...
[16:25:01] Judge: I am from Dorset...where are you from?
[16:25:08] Entity: I am from Ukraine, from the city called Odessa. You might have heard about it.

Some Actual Conclusions, or Lessons, or At Least Conjectures

I am writing this essay because I am puzzled.

—rpg, “in the control room of the banquet,” 2016 [31]

When I wrote this, I really meant it. Stephen Dobyns wrote:

The process of writing is a process of discovery. One never begins writing knowing what the end product will be. It is found along the way.

—Dobyns, *Best Words, Best Order*, 1996 [32]

I started this essay with vague questions; after working through the writing, some questions got sharper, I formed some hypotheses, and I sowed a set of slowly dawning insights.

1. The Turing test has two parts: the imitation game part where dodging the question is ok, and the consciousness-seeking part where the interrogator tries hard to find evidence of a consciousness—a consciousness with a point of view.

2. I went to the writers’ conference believing my haiku fib would be easily sniffed out, but it wasn’t. InkWell was that good at writing haiku.

3. Writing poetry—especially haiku—can be viewed as balancing a set of constraints and influences, some of which are about sense, others about tone, and still others about form, music, and emphasis. When these are specified as positive forces, good but conservative poems emerge; when they are negative, wild, imagistic, or even surreal poems do. When anti-sense is adopted as a goal alongside strong music, the result can be beautiful and strange new language—what poets call *dissociative* writing.

4. Producing good ordinary prose is harder than producing poetry because the information transfer part of discursive prose requires deep use of logic and a strong grasp on sense.

5. Producing short pieces of text—poetry or prose—requires a way to reason explicitly about the meanings and senses of words, phrases, clauses, and sentences. A purely Monte-Carlo-like or statistical approach misses nuance too often.

6. The mechanisms I put into InkWell for debugging purposes—the *representational content*—turned out to be a first step toward a self-model that (transparently) includes the self. These are components of consciousness as theorized by Thomas Metzinger [17]. These mechanisms include domain-specific languages for directing an artistic optimization-based, metaphorical writing process (InkWell uses pre-formed fragments in these languages and also creates poetry-writing programs in them); source code calculated for the optimization engine and force strengths that source code aims at; intermediate and discarded analyses; explicit language-use models including personality analysis, which are used in a write / read / evaluate process (much as human writers do); and an evolving, remembered language-use model for InkWell itself so it

can know what it has said before, and how it was said—this is used to steer its future artistic choices.

7. At the outset of the research, I conjectured that language-only representations would do. That is, I wanted to see how far I / InkWell could get by reasoning only about words, phrases, and sentences alongside some personality models and ngram-based language-use models. There are no externally provided ontologies. Everything InkWell knows about dogs, for example, comes from WordNet, its glosses, its example usages, and its synonym types. There are no common-sense reasoning or Monte-Carlo-like components. I felt the conjecture was valid, and the writers’ conference results reinforced that feeling.

8. I invented some mechanisms for reasoning about the senses and meanings of words and phrases along with very similar ways for the senses and meanings of contextual texts (haloes, for example) to influence word choice for purposes of subtext, mood, connotation, tone, etc. These turned out to work better than expected.

9. From CG (the workshop poet) I learned that poets (and others) in seeking evidence of consciousness and the inner lives of others rely not only on the words on the page, but on a ratcheting process wherein *some* evidence bootstraps *more* evidence, and on other direct hints—such as the person sitting right in front of them. I believed—and argued at the conference—that only the words on the page matter. But the poets in my workshop worked hard to find evidence of a human in the haiku, and they frequently seemed to work me directly into their investigations. I saw hints of this tenacious probing in Turing’s essay and especially in the Pickwick Test.

10. Working on InkWell is not like working on other software development projects. It is like writing an essay or a poem. To paraphrase Dobyns: “The process of programming is a process of discovery. One never begins coding knowing what the end product will be. It is found along the way.” From the dozen or so tries at sense making to the five (five!) different approaches to algorithmic rhyming, I am constantly revising my understanding.

11. Machine learning is hot. Turing anticipated it to a modest extent. Symbolic AI seems dead. The least well supported conjecture to come out of this is that consciousness requires an effective but transparent self-model that includes itself in it, and that such a self-model seems to require something more like symbolic AI than (only) machine-learned neural nets, classifiers, and sensors. Deep learning produces “processing units” and layers that are like abstractions, but these would need to be further abstracted, hand manipulated, or “learned upon” to be able to observe and explain intentionality (to act the way Searle requires consciousness to act). It seems likely that symbolic AI married to machine-learned AI is the best route to that right now.