Consciousness & Mood-Influenced Processing

Richard P. Gabriel
Hasso-Plattner Institute
3636 Altamont Way
Redwood City CA 94062
rpg@dreamsongs.com

Abstract

I believe there are two things we should consider for inclusion in the working Standard Model of the Mind—or at least for inclusion in the discussion. One is a place for consciousness, and the other is a set of mechanisms for mood, emotion, and general cognitive influence.

In July 2015 I took eighteen haiku-like poems to a writers' conference and presented them as my own work.[1] In reality, a program I created called "InkWell" wrote them, and I intended to execute a very informal variant of the Turing Test using the intense writers' workshop process (Gabriel 2002). I was not trying to execute a proper, scientifically valid Turing Test, but just trying to get a feel for how trained poets would respond to InkWell's work.

∼

In the Winter of 2014 I programmed InkWell, my English language revision system (Gabriel, Chen, and Nichols 2015) (Gabriel 2016) (Gabriel 2014), to write haiku—just to see whether it could do so plausibly. I let the system run overnight generating about 2,000 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 around ten haiku templates[2] as starting points—these essentially act as metaphor structures by setting a comparison starting point. And those four haiku are good—not just human-like, but good poetry with two of them close to being exceptional.

InkWell was designed as a creative writer's assistant, and my research programme was to explore *writerly thinking* rather than *information transfer*.[3] InkWell "knows" a lot about words, personality, sentiment, word noise, rhythm, connotations, and writing. Its vocabulary is about five times larger than the American average. 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 influences, 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.

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

∼

I worked 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 setting: a writers' workshop with three other trained and well-published poets.

∼

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" (Turing 1950). 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.

In his discussion of how the imitation game might go, 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

[1] I hold an MFA in Creative Writing (Poetry), and have published a small book of poetry.

[2] There are current seventy such templates, written in a domain-specific language for poetry.

[3] This work was done at IBM Research.

∼

The experience at the writers' conference and later research into the Turing test taught me it has three major thrusts:

- understanding natural language
- cleverly avoiding weak topics
- discovering whether the test subject exhibits consciousness

In an essay published right before Turing's, Geoffrey Jefferson wrote the following (Jefferson 1949):

> *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

One of the poets in the workshop said the following:

> 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; music seems important. Some are observations, some are moments, some are philosophical and very large (and not just the words, but the ideas). –DC

Another poet said this:

> 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

DC also said this:

> 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—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

From the workshop poets 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.

Turing himself wrote this in his essay:

> *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. Turing 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.[4] 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. 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. I believe I could program InkWell to access more gently this self model 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 n-grams 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.

∼

Thomas Metzinger (Metzinger 2010) and many others have been studying the nature of consciousness, along with mechanisms that support it. Consciousness is mostly the appearance of a world to a being; it is an awareness rather than a reaction. A thing that moves away from something hot is different from a thing that recognizes that thing as being hot. To be more precise, one of the current theories is that consciousness is a model of reality, generally within a being, and the most elaborate known levels of consciousness includes a model of the self. This model is built of neurons in people, and can be built of data, data structures, control structures, and relationships in software and computers. This implies that reality contains within it a model of itself, via the beings (and artifacts) with consciousness.

[4] This is called the "Pickwick" test, because Turing's essay describes a series of questions about Charles Dickens's "The Pickwick Papers."

> *The conscious brain is a biological machine—a reality engine—that purports to tell us what exists and what doesn't. It is unsettling to discover that there are no colors out there in front of your eyes. The apricot pink of the setting sun is not a property of the evening sky; it is a property of the internal model of the evening sky, a model created by your brain. The evening sky is colorless.*
>
> —Metzinger, *The Ego Tunnel*, 2010

Metzinger calls the consciousness a *phenomenal self-model (PSM)*. It is necessarily low-resolution because the world is rich and our sensory apparatus limited. Similarly, our sensory apparatus is limited in how well it can sense the brain itself.

He says of it:

> *The PSM of Homo Sapiens is probably one of nature's best inventions. It is an efficient way to allow a biological organism to consciously conceive of itself (and others) as a whole. Thus it enables the organism to interact with its internal world as well as with the external environment in an intelligence and holistic manner.*
>
> —Metzinger, *The Ego Tunnel*, 2010

∼

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 I "talk" to InkWell about these inner changes and factors, I do so in a nonhuman language, and InkWell responds in the same language.

If I were to try to program InkWell to respond to questions more naturally than inspecting complicated internal data structures, I would consider creating such a low-resolution model that would keep track of what InkWell was doing from minute to minute, and in a representation that was suitable for explaining what InkWell was thinking—as best it can. For symbolic parts of InkWell these representations would likely be simpler to design than ones that represent machine-learned observations etc. This model would be InkWell's consciousness.

∼

I propose a place in the Standard Model for a phenomenal self-model—for a consciousness.

∼

The second proposal is harder to explain. InkWell uses a variety of material to influence word and phrase choices beyond what's needed for information transfer—mood, emotion, and general cognitive influence:

- musical devices like Rhyme, Echo, and Rhythm
- conformance to general n-grams
- conformance to writer-specific n-grams
- deviation from self n-grams when seeking novelty
- OCEAN personality traits
- semantic "senses" based on a generalized word2vec-like representation, built according to a variety of algorithms depending on the purpose
- conformance to other semantic senses (halos, for example)

Moreover, conformance can take the form of nonconformance. Some of these have a significant machine-learned component. For example, the algorithm for creating sense structures from expanded WordNet entries was tuned using a genetic algorithm, and an algorithm for guessing phonetic spelling was created using machine learning.

These are called "mood-influencers" because they are used to modulate the tone and word choices InkWell pursues. These factors are over and above the concern of presenting the "intended" meaning precisely and accurately. Information transfer might be the most important concern for many written texts, but not for all of them.

One of the original goals for InkWell was to serve as a revision engine, which would take a passage of text and recast it to "sound like" another person.

Here is a simple example of influencing word and phrase choice based on non-information-transfer concerns. 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 word and phrase choice. For example, here is the first line of the last stanza of Robert Frost's "Stopping by Woods on a Snowy Evening":

> *The woods are lovely, dark, and deep*

This already is more than information transfer: "The woods are lovely" is innocuous enough, but the word "dark" implies something mysterious and even sinister or malevolent. Then the word "deep" tells us it's a place that can be explored for a long time, a place to get lost in, to lose one's self in. Thus the woods are a trap, and the speaker acknowledges that were it not for life needing to be lived, he or she could easily be tempted to stop here, maybe forever.

Holding every other influence fixed, Inkwell might revise it to this

> *The woods are glorious, not too light, and not too shallow*

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 not very ugly, black, and heavy*

when given this halo:

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

Notice that InkWell uses the locution "not too *adj*." Even though there is semantic opposition in "not too," the reader still experiences the happiness-hinting word "light" and the less ominous "shallow" in the first revision; similarly for "ugly" in the anger-influenced revision.

The point is that auxiliary mood influencers can make a significant difference to what is generated, and such mechanisms can be used to introduce subtexts, context, mood, emotion, personality, and style into what the software creates.

There are around thirty such knobs in InkWell, and several of them have a defined algebra for combining them. It was through the use of such an algebra that the following haiku was written in response to the request “Please write me a haiku on the subject of blues, guitar, and loud music”:[5]

tuned adrenalin
my music,
a beat-boogied headful

∼

I propose that the Standard Model be augmented with mechanisms that support mood-influenced processing.

Remarks On the Standard Model

Aspects of the existing Standard Model (SM) appear, at first, to be amenable to supporting forms of consciousness and mood influence, but my feeling is that there are some mismatches. For example, working memory coupled with procedural memory, being rule-based and pattern-directed over working memory, seem like a fine place for consciousness to reside. Perhaps it can, but I suspect the representations in working memory and perhaps the detailed nature of the rule-pattern language will prove too fine grained for the level of processing that happens in consciousness. Further, consciousness needs to include some sort of outside-world model, which implies a coarse description.

This observation leads to the following broad suggestions. First, we should think in terms of *descriptions* and not representations. That is, consciousness strives to *describe* what is in the world and also in the rest of the cognitive apparatus, not to represent those things fully and precisely. I suspect that a good description language needs to be “shallow” in some sense, as well as approximate. This does not necessarily require descriptions to be terse or compact.

Second, the descriptions consciousness maintains should be based on observations made by the consciousness machinery and not by various parts of the SM pushing descriptions into the consciousness model. Clearly the world doesn’t push descriptions into the mind; rather, the mind mines them.

Third, consciousness needs to be *effective*, which means that it should be able to initiate cognitive action in the rest of the SM, including revisions to memory. Such revisions are already anticipated by the SM, especially as a result of learning. However, being effective does not necessarily require immediate action—perhaps something more like turning a large ocean-going tanker or persuading a distracted dog to sit.

Fourth, consciousness needs to be able to perform cognitive-like processing, such as (grossly) simulating planned actions in the real world.

∼

A mood influencer acts like a hormone or a law of physics—it is pervasive: not obviously associated with any single object, but with processes that select, alter, or interact with objects. In InkWell, things like halos are injected explicitly into selection processes for words, phrases, and linguistic structures, and the result is as if InkWell were in a particular mood when writing a passage.

Pervasiveness is not exactly like attaching behavior to every physical object. In the real world, gravity can never be accidentally left out of a particular physical object; similarly, a mood influencer should not be a mechanism that needs to be explicitly attached to all “modules” or “objects.”

How to do this in the SM?

One way is for mood influencers to operate as if within the “interpreter” that executes the SM machinery—thus mood influencers would be able to stick their noses in wherever they please. An approximation to this would be to attach mood influencers to the pattern-directed rule invocation mechanism.

Another way to inject mood influencers might be to attach them to the consciousness mechanism. To make this work the consciousness mechanism would need to act a little like the way electric motors do in hybrid automobiles: in many cases the regular SM mechanisms would work by themselves, but sometimes the consciousness mechanism would kick in to help out, and other times the consciousness mechanism would act by itself. This requires consciousness to be (possibly only weakly) effective, but it also implies that mood influencers are shallow and only approximate.

The Haiku

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

the powerful head
designates its powerful head
to support cognition

this grave—
no one sees it
mortality, mortality

a bitch,
this deep in trick
a fortiori not a man

not this fatalist murderousness,
deathwatch,
but your dead subroutine

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

[5] InkWell almost always generates grammatical haiku (prepositions are sometimes off), but not always haiku meaningful to people. As with human poets, selection and sometimes further revision needs to be done. You should walk away with the impression that InkWell is intriguing, but not with the impression that it is a reliably good poet. None of the haiku in this essay were revised.

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

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

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

 a crooked rag day—
by myself
 dunking distracted sardines

 an on-the-far-side summer night—
whipping up high tea,
 we stripped pickles

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

 a reasonable assumption—
by myself,
  sampling in chocolate

 rural signal,
cannot understand Oregon
 —agricultural

parted in the middle—
 the authority of the air conditioner
  perfection in the brightness

the hostile defense
 leads its problematic rear,
  the rear of frustration

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

References

Gabriel, R. P.; Chen, J.; and Nichols, J. 2015. Inkwell: A creative writer’s creative assistant. *Creativity & Cognition*.

Gabriel, R. P. 2002. *Writers’ Workshops & the Work of Making Things: Patterns, Poetry. . . .* New York: Pearson.

Gabriel, R. P. 2014. I throw itching powder at tulips. *Onward! Essays*.

Gabriel, R. P. 2016. in the control room of the banquet. *Onward! Essays*.

Jefferson, G. 1949. The mind of mechanical man. *British Medical Journal*.

Metzinger, T. 2010. *The Ego Tunnel: The Science of the Mind and the Myth of the Self.* Basic Books.

Turing, A. M. 1950. Computing machinery and intelligence. *Mind*.