Is It Possible to Build Dramatically Compelling Interactive Digital Entertainment

(in the form, e.g., of computer games)?*

Selmer Bringsjord
The Minds & Machines Laboratory
Dept. of Philosophy, Psychology & Cognitive Science
Department of Computer Science
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180-3590 USA
selmer@rpi.edu • http://www.rpi.edu/~brings

2.16.01

Abstract

Lots of computer games are compelling. E.g., I find even current computerized poker games quite compelling, and I find *The Sims* downright fascinating; doubtless you have your own favorites. But our planet isn't graced by even one *dramatically compelling* computer game (or, more generally, one such interactive digital entertainment). The movie *T2*, Dante's *Inferno*, *Hamlet*, Gibson's prophetic *Neuromancer*, the plays of Ibsen — these things are dramatically compelling: they succeed in no small part because they offer captivating narrative, and all that that entails (e.g., engaging characters). There is no analogue in the interactive digital arena, alas. Massively multi-player online games are digital, interactive, and entertaining — but they have zero literary power (which explains why, though *T2* engages young kids through at least middle-aged professors, such games are demographically one-dimensional). The same can be said, by my lights, for all other electronic genres.

This state of affairs won't change unless a number of key challenges are conquered; and conquering them will require some seminal advances in the intersection of artificial intelligence (AI) and narrative. (E.g., since interactive digital narrative will need to be crafted and massaged *as the story is unfolding*, computers, not slow-by-comparison humans, will need to be enlisted as at least decent dramatists — but getting a computer to be a dramatist requires remarkable AI.) In this paper, I discuss one of these challenges for the start of the new millennium: the problem of building dramatically compelling virtual characters. Within this challenge I focus upon one property such characters presumably must have: viz., autonomy.

*I'm indebted to Dave Ferrucci, Devin Croak, and Marc Destefano.

1 The Issue

We have dramatically compelling *non*-interactive digital (= electronic) entertainment: sit down and watch *The Matrix* (see Figure 1) or *T2* (see Figure 2). We have compelling interactive digital entertainment; my favorites include console-based sports games.[1] We have dramatically compelling interactive entertainment; for example, improvizational theatre. What we *don't* have is dramatically compelling interactive digital entertainment. People sometimes tell me that there is a counter-example to my negative view to be found in online multi-player games, but though such games are digital, interactive, and entertaining — they have zero literary power (which explains why, though *T2* engages young kids through at least middle-aged professors, such games are demographically one-dimensional). The same can be said, by my lights, for all other electronic genres. Can we build systems that imply and affirmative answer to the title of this paper? My answer is: "Maybe, maybe not; but at any rate I can tell you, at least in broad strokes, what some of the hurdles are, from the standpoint of AI. And I can tell you, today, about one specific hurdle."

Figure 1: From a Dramatically Compelling Scene in *The Matrix*

2 Realism About Narrative and AI

Before we go any further, let's make sure we're realistic about the driving question, and thereby start with a provisional answer of "I don't know." Such realism will buck the trend. For unfortunately, realistic positions on the advance of AI are rather hard to come by. There are two reasons for this.

The first reason is that lots of people are either ignorant of or tendentiously choose to ignore the underlying mathematical facts. These facts include that some problems can be solved by computing machines, and others can't, and that most can't. So whenever you ask whether a problem P can be solved by a computing machine, where P is such that it isn't known that there is an algorithm for solving it, honesty should imply wait-and-see agnosticism.[2]

The second reason why realism in the face of questions like that which drives the present investigation is in short supply is that we have lots of silly prophets. For example, in the June 19, 2000 issue of TIME magazine, devoted to "The Future of Technology," we hear from author

[1] Playing soccer against someone with Playstation 2 running to a large plasma display is, for me, fun. Perhaps you, on the other hand, prefer a current online multi-player game or two.

[2] I have recently argued that deciding whether some story is interesting is a computationally unsolvable problem. See "Chapter 5: The Narrative-Based Refutation of Church's Thesis" in (Bringsjord & Ferrucci 2000).

Figure 2: POV of the Terminator in *T2*

and inventor Ray Kurzweil that nanobots (microscopic robots) will by 2030 be able to map out a synthetic duplicate of your brain after you swallow (yes, swallow) a few of them. This duplicate will be instantiated in computer hardware 10 million times faster than the sluggish, old-fashioned grey stuff inside your cranium; the result will be an artificial intelligence immeasurably more clever than you. Vernor Vinge, associate professor of mathematics and computer science at San Diego State University, is another example. Prophesying for the *Chronicle of Higher Education* (July 12, 2000; online edition), he gives us a more compressed timeline: by his lights, on the strength of the trend that the speed of computer hardware doubles every 18 months, computers will be more intelligent than all humans at some point within 20 years. This point he calls "The Singularity," which ushers in post-humanity, an age in which humans are left in the dust by machines that get exponentially smarter by the day (if not the nanosecond). For a third example, consider Hans Moravec, who in his latest book, *Robot: Mere Machine to Transcendent Mind*, informs us that because hardware is getting faster at the rate Vinge cites, by 2040 "fourth generation" robots will exceed humans in all respects, from running companies to writing novels. Such robots will evolve to such lofty cognitive heights that we will stand to them as single-cell organisms stand to us today. Many others in the field of Artificial Intelligence (AI) predict the same sensational future unfolding on about the same rapid schedule.

The gaming industry will not be this lucky; I'm sure of it. Today I'll tell you, briefly, why.

3 One Presupposition: There's No Free Lunch

Let make explicit one presupposition before we begin in earnest. I assume that AI in general, along with the AI part of the gaming industry in particular, have realized that non-logicist AI isn't magic. What do I mean? I mean that people should be smart enough now to concede that no system is going to automatically learn, through subsymbolic, numerical processing, how to, say (and this is the challenge I'm going to focus on below), assemble robust, autonomous characters in a game. Automated learning through artificial neural networks and genetic algorithms are fine for certain applications, but gone, I assume, are the days of wild optimism about the ability of such learning techniques to automatically yield (to stick with this example) full-blooded NPCs.[3] We really only have two overall approaches to AI: one based on logic, and one based on subsymbolic processing. This paper is written from the perspective of logic. If you think that you have a way of solving the problems I'm talking about without using logic, I wish you well.

[3]The connectionist-logicist clash in AI is discussed in: (Bringsjord & Ferrucci 2000, Bringsjord 1991, Bringsjord & Ferrucci 1998). The last of these publications constitutes an introduction to logicist AI.

4 A List of Some Challenges

Very well. So, why is it that building dramatically compelling interactive digital entertainment (in the form, e.g., of games) is so difficult? There are many reasons, among which fall the following.

C1: Formalizing Literary Themes. If Dave Ferrucci and I are right, plotlines and so-called 3-dimensional characters aren't enough for a computer to generate first-rate narrative: you also need to instantiate immemorial themes — betrayal (the one we focus on in (Bringsjord & Ferrucci 2000)), self-deception (a theme we employ in BRUTUS), unrequited love, revenge, and so on. If such themes are to be used by story-managing machines, they must be represented; if they are to be represented and exploited, they need to be *rigorously* represented and reasoned over. Such rigorous representation and reasoning is very hard to come by.

C2: Story Mastery. After interactive drama begins, things can devolve toward the uninteresting. If "hack-and-slash" is all that is sought from an interactive game, then such devolution may be acceptable. But if genuine drama is desired, then something or someone must ensure that what happens is dramatically interesting. One possibility (with respect to a multi-player online game) is to have a human or humans oversee the action as it unfolds, and make changes that keep things "on track." For obvious reasons, in a rapidly progressing game with thousands of human players, this is impracticable; the possibility of human oversight is purely conceptual. So we must turn to computers to automate the process. But how? How is the automation to work? I've assumed that if a program could be built that writes compelling fiction, we might thereby have taken significant steps toward a program that can serve as a story master.

C3: Building Robust, Autonomous Characters. A *sine qua non* for compelling literature and drama is the presence of robust, autonomous, and doxastically sophisticated[4] characters. In short, such literature and drama exploits the central properties of being a person. (In many cases, great stories come to be remembered in terms of great characters.) This presents a problem for interactive electronic entertainment: how do we build an electronic character that has those attributes that are central to personhood, and whose interaction with those humans who enter the virtual worlds is thereby compelling?

C4: Personalization. If virtual characters are going to react intelligently to you as user or gamer, they must, in some sense, *understand* you. This problem is directly related to C3, because the characters must have sophisticated beliefs about you and your beliefs, etc.

In the remainder of this paper, I focus on C3, and moreover I focus within this challenge on the specific problem of building autonomous characters. The plan is as follows. I begin by reviewing the concept of an **intelligent agent** in AI. I then explain the clash between this limited concept and the kind of properties that are distinctive of personhood; one of these properties is autonomy, or "free will." In order to highlight the problem of imparting autonomy to a virtual character, I turn to what I have dubbed "The Lovelace Test." I conclude with a disturbing argument that seems to show that virtual characters, as intelligent agents, can't be autonomous, because they would inevitably fail this test. I do intimate my own reaction to this argument.

5 Intelligent Agents

As the century turns, all of AI has been to an astonishing degree unified around the conception of an intelligent agent. The unification has in large part come courtesy of a comprehensive textbook intended to cover literally *all* of AI: Russell and Norvig's (1994) *Artificial Intelligence: A Modern Approach* (*AIMA*), the cover of which also displays the phrase "The Intelligent Agent Book." The

[4] A character is doxastically sophisticated if it can reason over its beliefs about the beliefs other characters and human users have about beliefs, etc.

Table 1: Lookup Table for TABLE-DRIVEN-AGENT

Percept	Action
001	"Red"
010	"Green"
100	"Blue"
011	"Yellow"
111	"Black"

overall, informal architecture for an intelligent agent is shown in Figure 3; this is taken directly from the *AIMA* text. According to this architecture, agents take percepts from the environment, process them in some way that prescribes actions, perform these actions, take in new percepts, and continue in the cycle.[5]

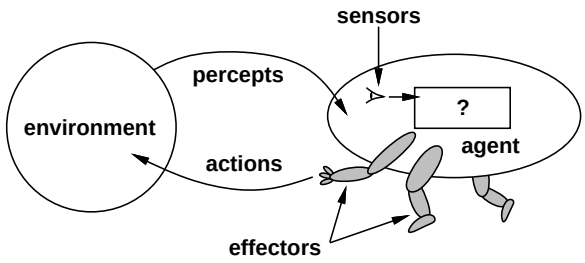

Figure 3: The Architecture of an Intelligent Agent

In *AIMA*, intelligent agents fall on a spectrum from least intelligent to more intelligent to most intelligent. The least intelligent artificial agent is a "TABLE-DRIVEN-AGENT," the program (in pseudo-code) for which is shown in Figure 4. Suppose that we have a set of actions each one of which is the utterance of a color name ("Green," "Red," etc.); and suppose that percepts are digital expressions of the color of an object taken in by the sensor of a table-driven agent. Then given Table 1 our simple intelligent agent, running the program in Figure 4, will utter (through a voice synthesizer, assume) "Blue" if its sensor detects 100. Of course, this is a stunningly dim agent. What are smarter ones like?

In *AIMA* we reach artificial agents that might strike some as rather smart when we reach the level of a "knowledge-based" agent. The program for such an agent is shown in Figure 5. This

[5]The cycle here is strikingly similar to the overall architecture of cognition described by Pollock (1995).

```
function TABLE-DRIVEN-AGENT(percept) returns action
  static: percepts, a sequence, initially empty
          table, a table, indexed by percept sequences, initially fully specified

  append percept to the end of percepts
  action ← LOOKUP(percepts, table)
  return action
```

Figure 4: The Least Intelligent Artificial Agent

```
function KB-AGENT(percept) returns an action
  static: KB, a knowledge base
          t, a counter, initially 0, indicating time

  TELL(KB, MAKE-PERCEPT-SENTENCE(percept, t))
  action ← ASK(KB, MAKE-ACTION-QUERY(t))
  TELL(KB, MAKE-ACTION-SENTENCE(action, t))
  t ← t + 1
  return action
```

Figure 5: Program for a Generic Knowledge-Based Agent

program presupposes an agent that has a knowledge-base (KB) in which what the agent knows is stored in formulae in the propositional calculus, and the functions

- TELL, which injects sentences (representing facts) into KB;
- MAKE-PERCEPT-SENTENCE, which generates a propositional calculus sentence from a percept and the time t at which it is experienced; and
- MAKE-ACTION-SENTENCE, which generates a declarative fact (in, again, the propositional calculus) expressing that an action has been taken at some time t

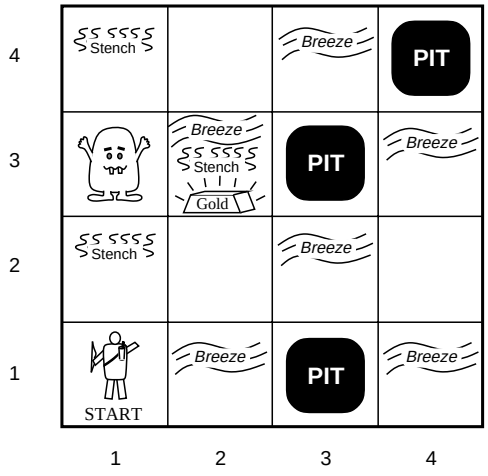

Figure 6: A Typical Wumpus World

which give the agent the capacity to manipulate information in accordance with the propositional calculus. (One step up from such an agent would be a knowledge-based agent able to represent and reason over information expressed in full first-order logic.) A colorful example of such an agent is one clever enough to negotiate the so-called "wumpus world." An example of such a world is shown in Figure 6. The objective of the agent that finds itself in this world is to find the gold and bring it back without getting killed. As Figure 6 indicates, pits are always surrounded on three sides by breezes, the wumpus is always surrounded on three sides by a stench, and the gold glitters in the square in which it's positioned. The agent dies if it enters a square with a pit in it (interpreted as falling into a pit) or a wumpus in it (interpreted as succumbing to an attack by the wumpus). The percepts for the agent can be given in the form of quadruples. For example,

(Stench,Breeze,Glitter,None)

means that the agent, in the square in which it's located, perceives a stench, a breeze, a glitter, and no scream. A scream occurs when the agent shoots an arrow that kills the wumpus. There are a number of other details involved, but this is enough to demonstrate how command over the propositional calculus can give an agent a level of intelligence that will allow it to succeed in the wumpus world. For the demonstration, let $S_{i,j}$ represent the fact that there is a stench in column i row j, let $B_{i,j}$ denote that there is a breeze in column i row j, and let $W_{i,j}$ denote that there is a wumpus in column i row j. Suppose now that an agent has the following 5 facts in its *KB*.

1. $\neg S_{1,1} \wedge \neg S_{2,1} \wedge S_{1,2} \wedge \neg B_{1,1} \wedge B_{2,1} \wedge \neg B_{1,2}$
2. $\neg S_{1,1} \rightarrow (\neg W_{1,1} \wedge \neg W_{1,2} \wedge \neg W_{2,1})$
3. $\neg S_{2,1} \rightarrow (\neg W_{1,1} \wedge \neg W_{2,1} \wedge \neg W_{2,2} \wedge \neg W_{3,1})$
4. $\neg S_{1,2} \rightarrow (\neg W_{1,1} \wedge \neg W_{1,2} \wedge \neg W_{2,2} \wedge \neg W_{1,3})$
5. $S_{1,2} \rightarrow (W_{1,3} \vee W_{1,2} \wedge W_{2,2} \wedge W_{1,3})$

Then in light of the fact that

$$\{1, \ldots, 5\} \vdash W_{1,3}$$

in the propositional calculus,[6] the agent can come to know (= come to include in its *KB*) that the wumpus is at location column 1 row 3 — and this sort of knowledge should directly contribute to the agent's success.

In my lab, a number of students have built actual wumpus-world-winning robots; for a picture of one toiling in this world see Figure 7.

Now I have no problem believing that the techniques and formalisms that constitute the agent-based approach preached in *AIMA* are sufficient to allow for the construction of characters that operate at the level of animals. But when we reach the level of personhood, all bets, by my lights, are off.

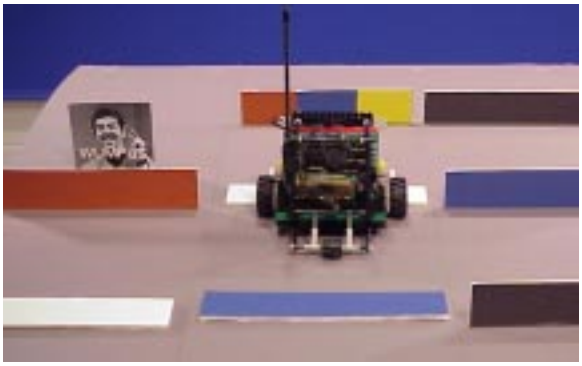

Figure 7: A Real-Life Wumpus-World-Winning Robot in the Minds & Machines Laboratory *(Observant readers may note that the wumpus here is represented by a figurine upon which appears the (modified) face of the Director of the M&M Lab: Bringsjord.)*

6 The Roadblock: Personhood

Why is it that intelligent agent techniques will allow us to build virtual rats, parrot, and chimps, but fail when we attempt to build virtual persons? They will fail because intelligent agent architectures, formalisms, tools, and so on are impotent in the face of the properties that distinguish

[6] The proof is left to sedulous readers.

persons. What are these properties? Many philosophers have taken up the challenge of answering this question, but for present purposes it suffices to call upon an account of personhood offered in (Bringsjord 1997); in fact, it suffices to list here only five of the properties offered in that account, viz.,[7]

1. ability to communicate in a language
2. autonomy ("free will")
3. creativity
4. phenomenal consciousness
5. robust abstract reasoning (e.g., ability to create conceptual schemes, and to switch from one to another)

For the sake of argument I'm prepared to follow Turing and hold that AI will engineer not only the communicative powers of a parrot and a chimp, but also the linguistic powers of a human person. (This concession requires considerable optimism: The current state-of-the art in AI is unable to create a device with the linguistic capacity of a toddler.) However, it's exceedingly hard to see how each of the four remaining properties can be reduced to the machinery of the intelligent agent paradigm in AI. Intelligent agents don't seem to originate anything; they seem to do just what they have been designed to do. And so it's hard to see how they can originate decisions and actions ("free will") or artifacts (creativity). At least at present, it's hard to see how phenomenal consciousness can be captured in any third-person scheme whatever (and as many readers will know, a number of philosophers — Nagel, e.g. — have argued that such consciousness can never be captured in such a scheme), let alone in something as austere as what AI engineers work with. And those in AI who seek to model abstract reasoning know well that we have only begun to show how sophisticated abstract reasoning can be cast in well-understood computable logics. For all we know at present, it may be that some of this reasoning is beyond the reach of computation. Certainly such reasoning cannot be cashed out in the vocabulary of *AIMA*, which stays firmly within extensional first-order logic.

But let's focus, as I said I would, on the issue of autonomous intelligent agents. I believe I have a way of sharpening the challenge that this issue presents to those who aspire to create dramatically compelling interactive electronic entertainment. This way involves subjecting would-be autonomous virtual characters to a form of the Lovelace Test. But first, I have to introduce the test.

7 The Lovelace Test

As you probably know, Turing predicted in his famous "Computing Machinery and Intelligence" (1964) that by the turn of the century computers would be so smart that when talking to them from a distance (via email, if you will) we would not be able to tell them from humans: they would be able to pass what is now known as the Turing Test (TT). Well, New Year's Eve of 1999 has come and gone, all the celebratory pyrotechnics have died, and the fact is: AI hasn't managed to produce a computer with the conversational punch of a toddler.

But the really depressing thing is that though progress toward Turing's dream is being made, it's coming only on the strength of clever but shallow trickery. For example, the human creators of artificial agents that compete in present-day versions of TT know all too well that they have merely tried to *fool* those people who interact with their agents into believing that these agents

[7]The account is streamlined in the interests of space. For example, because people sleep (and because they can be hypnotized, etc.), a person would be a creature with the *capacity* to have properties like those listed here.

really have minds. In such scenarios it's really the human creators against the human judges; the intervening computation is in many ways simply along for the ride.

It seems to me that a better test is one that insists on a certain restrictive epistemic relation between a an artificial agent A, its output o, and the human architect H of S — a relation which, roughly speaking, obtains when H cannot account for how A produced o. I call this test the "Lovelace Test" in honor of Lady Lovelace, who believed that only when computers *originate* things should they be believed to have minds.

7.1 The Lovelace Test in More Detail

To begin to see how LT works, we start with a scenario that is close to home for Bringsjord and Ferrucci, given their sustained efforts to build story generation agents: Assume that Jones, a human AInik, attempts to build an artificial computational agent A that doesn't engage in conversation, but rather creates stories — creates in the Lovelacean sense that this system *originates* stories. Assume that Jones activates A and that a stunningly belletristic story o is produced. We claim that if Jones cannot explain how o was gemerated by A, *and* if Jones has no reason whatever to believe that A succeeded on the strength of a fluke hardware error, etc. (which entails that A can produce other equally impressive stories), then A should at least provisionally be regarded genuinely creative. An artificial computational agent passes LT if and only if it stands to its creator as A stands to Jones.

LT relies on the special epistemic relationship that exists between Jones and A. But 'Jones,' like 'A,' is of course just an uninformative variable standing in for any human system designer. This yields the following rough-and-ready definition.

Def$_{LT}$ 1 Artificial agent A, designed by H, passes LT if and only if

1 A outputs o;

2 A's outputting o is not the result of a fluke hardware error, but rather the result of processes A can repeat;

3 H (or someone who knows what H knows, and has H's resources[8]) cannot explain how A produced o.

Notice that LT is actually what might be called a *meta*-test. The idea is that this scheme can be deployed for any partcular domain. If conversation is the kind of behavior wanted, then merely stipulate that o is an English sentence (or sequence of such sentences) in the context of a converation (as in, of course, TT). If the production of a mathematical proof with respect to a given conjecture is what's desired, then we merely set o to a proof. In light of this, we can focus LT on the particular kind of interaction appropriate for the digital entertainment involved.

Obvious questions arise at this point. Three are:

Q1 What resources and knowledge does H have at his or her disposal?

Q2 What sort of thing would count as a successful explanation?

Q3 How long does H have to cook up the explanation?

The answer to the third question is easy: H can have as long as he or she likes, within reason. The proffered explanation doesn't have to come immediately: H can take a month, months, even a year or two. Anything longer than a couple of years strikes us as perhaps unreasonable. We realize

[8]For example, the substitute for H might be a scientist who watched and assimilated what the designers and builders of A did every step along the way.

that these temporal parameters aren't exactly precise, but then again we should not be held to standards higher than those pressed against Turing and those who promote his test and variants thereof.[9] The general point, obviously, is that H should have more than ample time to sort things out.

But what about Q1 and Q2? The answer to Q1 is that H is assumed to have at her disposal knowledge of the architecture of the agent in question, knowledge of the KB of the agent, knowledge of how the main functions in the agent are implemented (e.g., how TELL and ASK are implemented), and so on (recall the summary of intelligent agents above). H is also assumed to have resources sufficient to pin down these elements, to "freeze" them and inspect them, and so on. I confess that this isn't exactly precise. To clarify things, I offer an example. This example is also designed to provide an answer to Q2.

To fix the context for the example, suppose that the output from our artificial agent A' is a resolution-based proof which settles a problem which human mathematicians and logicians have grappled unsuccessfully with for decades. This problem, suppose, is to determine whether or not some formula ϕ can be derived from some (consistent) axiom set Γ. Imagine that after many years of fruitless deliberation, a human H' encodes Γ and $\neg\phi$ and gives both to OTTER (a well-known theorem prover; it's discussed in (Bringsjord & Ferrucci 2000)), and OTTER produces a proof showing that this encoding is inconsistent, which establishes $\Gamma \vdash \phi$, and leads to an explosion of commentary in the media about "brilliant" and "creative" machines, and so on.[10] In this case, A' doesn't pass LT. This is true because H, knowing the KB, architecture, and central functions of A' will be able to give a perfect explanation for the behavior in question. I routinely give explanations of this sort. The KB is simply the encoding of $\Gamma \cup \{\phi\}$, the architecture consists in the search algorithms used by OTTER, and the main functions consist in the rules of inference used in a resolution-based theorem prover.

Here, now, given the foregoing, is a better definition:

Def$_{LT}$ 2 Artificial agent A, designed by H, passes LT if and only if

1 A outputs o;

2 A's outputting o is not the result of a fluke hardware error, but rather the result of processes A can repeat;

3 H (or someone who knows what H knows, and has H's resources) cannot explain how A produced o by appeal to A's architecture, knowledge-base, and core functions.

7.2 How do Today's Systems Fare in the Lovelace Test?

Today's systems, even those designed to either be, or seem to be, autonomous, fail LT. These designers can imagine themselves generating the output in question by merely manipulating symbols in accordance with the knowledge bases, algorithms, and code in question. We give an example of this kind of failure, an example that falls rather close to home for me.

7.3 Why BRUTUS Fails the Lovelace Test

The BRUTUS system is designed to appear to be literarily creative to *others*. To put the point in the spirit of the Turing Test, BRUTUS reflects a multi-year attempt to build a system able to play

[9] In (Bringsjord 1995), Bringsjord refutes propositions associated with TT by assuming for the sake of argument that some reasonable parameters π have been established for this test. But Turing didn't specify π, and neither have his present-day defenders.

[10] For a "real life" counterpart, we have OTTER's settling the Robbins Problem, presented as an open question in (Wos 1996).

the short short story game, or S^3G for short (Bringsjord 1998). (See Figure 8 for a picture of S^3G.)

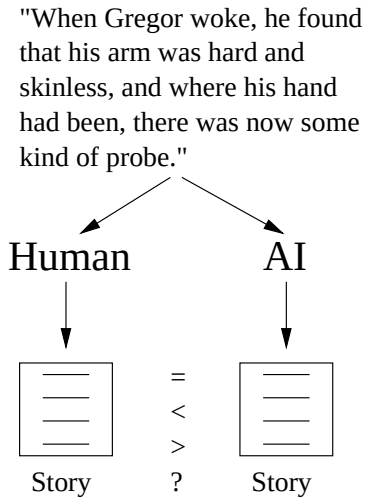

Figure 8: The Short Short Story Game, or S^3G for Short.

The idea behind S^3G is simple. A human and a computer compete against each other. Both receive one relatively simple sentence, say: "As Gregor Samsa awoke one morning from uneasy dreams he found himself transformed in his bed into a gigantic insect." (Kafka 1948, p. 67) Both mind and machine must now fashion a short short story (about 500 words) designed to be truly interesting; the more literary virtue, the better. The goal in building BRUTUS, then, is to build an artificial author able to compete with first-rate human authors in S^3G, much as Deep Blue went head to head with Kasparov.

How does BRUTUS fare? Relative to the goal of passing S^3G, not very well. On the other hand, BRUTUS can "author" some rather interesting stories (Bringsjord & Ferrucci 2000). Note that we have placed the term 'author' in scare quotes. Why? The reason is plain and simple, and takes us back to Lady Lovelace's objection: BRUTUS doesn't *originate* stories. He is capable of generating it because two humans, Bringsjord and Ferrucci, spent years figuring out how to formalize a generative capacity sufficient to produce this and other stories, and they then are able to implement part of this formalization so as to have a computer produce such prose. This method is known as *reverse engineering*. Obviously, with BRUTUS set to A and Bringsjord and Ferrucci set to H in the definition of LT, the result is that BRUTUS fails this test.

Let's now give you, briefly, a specific example to make this failure transparent. BRUTUS is programmed to produce stories that, are, at least to some degree, bizarre. The reason for this is that reader response research tells us that readers are engaged by bizarre material. Now, in BRUTUS, to express the bizarre, modifiers are linked with objects in frames named `bizzaro_modifiers`. Consider the following instance describing the `bizzaro` modifier `bleeding`.

```
instance bleeding is a bizzaro_modifier
    objects are {sun, plants, clothes, tombs, eyes}.
```

What Bringsjord and Ferrucci call **literary augmented grammars**, or just a LAGs, may be augmented with constraints to stimulate bizarre images in the mind of the reader. The following LAG for action analogies,

- `BizarreActionAnalogy` → `NP VP like ANP`

- `NP` → `noun_phrase`
- `ANP` → `modifier (isa bizzaro_modifier) noun (isa analog of NP)`

in conjunction with `bizzaro_modifiers`, can be used by BRUTUS to generate the following sentence.

Hart's eyes were like big bleeding suns.

Sentences like this in output from BRUTUS are therefore a function of work carried out by (in this case) Ferrucci. Such sentences do not result from BRUTUS thinking on its own.

8 The Conclusing Argument

What does the Lovelace Test buy us? What role does it play in connection with challenge C3? The overall idea is this. A truly autonomous virtual character is an intelligent agent that has those attributes constitutive of personhood, attributes that include autonomy. Operationalized, this means that truly autonomous virtual characters would pass LT. But such agents *can't* pass LT. Put in the form of an argument, and tied to the question that gives this paper its title, we have:

The Argument That Worries Me

1. Dramatically compelling interactive digital entertainment requires the presence in such entertainment of virtual persons, and therefore requires the presence of autonomous virtual characters.
2. Autonomous virtual characters would pass the Lovelace Test.
3. Autonomous virtual characters would be intelligent agents, in the technical sense of "intelligent agents" in use in AI (specifically in *AIMA*).
4. Intelligent agents fail the Lovelace Test.
5. ∴ Dramatically compelling interactive digital entertainment isn't possible.

What should the response be to this argument be? Perhaps you'll need to think about what your own reaction should be; the point of this paper is only to place the argument before you. Clearly, the argument is formally valid, that is, the logic is correct: 5. does follow from the premises. So to escape the argument, at least one of the premises must be rejected. My suspicion is that premise 1. is false, but that what's true is a relative, viz.,

1′. Dramatically compelling interactive digital entertainment requires the presence in such entertainment of *seemingly* autonomous virtual characters.

If this is right, those who design and build digital entertainment need, at bottom, to figure out ingenious ways of fooling players into believing that virtual characters are, in general, persons (and hence, among other things, autonomous). The job description for those intent on building dramatically compelling interative digital entertainment thus calls for those who can figure out the stimuli that impel gamers to believe they are interacting with virtual people, and then engineer a system to produce this stimuli in a principled way. This job description is decidely *not* filled by those in game development who have mastered the so-called present-day "art of character design," which is nicely summarized, e.g., in (Gard 2000). Why this is so, and what, as a practical engineering matter, needs to be done to extend present-day techniques — well, this will need to wait for another day.

References

Bringsjord, S. (1991), 'Is the connectionist-logicist clash one of ai's wonderful red herrings?', *Journal of Experimental & Theoretical AI* **3.4**, 319–349.

Bringsjord, S. (1995), Could, how could we tell if, and why should–androids have inner lives?, *in* K. Ford, C. Glymour & P. Hayes, eds, 'Android Epistemology', MIT Press, Cambridge, MA, pp. 93–122.

Bringsjord, S. (1997), *Abortion: A Dialogue*, Hackett, Indianapolis, IN.

Bringsjord, S. (1998), 'Chess is too easy', *Technology Review* **101**(2), 23–28.

Bringsjord, S. & Ferrucci, D. (1998), 'Logic and artificial intelligence: Divorced, still married, separated...?', *Minds and Machines* **8**, 273–308.

Bringsjord, S. & Ferrucci, D. (2000), *Artificial Intelligence and Literary Creativity: Inside the Mind of Brutus, a Storytelling Machine*, Lawrence Erlbaum, Mahwah, NJ.

Gard, T. (2000), 'Building character', *Game Developer Magazine* **7.5**, 28–37.

Kafka, F. (1948), The metamorphosis, *in* F. Kafka, t. W. Muir & E. Muir, eds, 'The Penal Colony', Schocken Books, New York, NY.

Pollock, J. (1995), *Cognitive Carpentry: A Blueprint for How to Build a Person*, MIT Press, Cambridge, MA.

Russell, S. & Norvig, P. (1994), *Artificial Intelligence: A Modern Approach*, Prentice Hall, Saddle River, NJ.

Turing, A. (1964), Computing machinery and intelligence, *in* A. R. Anderson, ed., 'Minds and Machines', Prentice-Hall, Englewood Cliffs, NJ, pp. 4–30.

Wos, L. (1996), *The Automation of Reasoning: An Experimenter's Notebook with* OTTER *Tutorial*, Academic Press, San Diego, CA.