Synthetic Worlds and Characters, and the Future of Creative Writing

Selmer Bringsjord[1] and **Alexander Bringsjord**[2]
Rensselaer AI & Reasoning (RAIR) Lab[1,2]
Department of Cognitive Science[1,2]
Department of Computer Science[1]
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA
draft of 3.17.09 2338 CO

The Future

Fenwick Gibsen ate breakfast with his wife each and every morning before commuting to his office. Suzannah Gibsen, Fenwick well knew, was a bit of a retro health nut; in fact, to say “a bit” was to understate the matter rather dramatically. She insisted on each of them consuming, side by side in their bay-windowed, east-facing breakfast room, at sunrise, a bowl filled with a mountain of crisp blueberries, under a layer of strawberries, atop a mound of oat brain suffused with an ample, underlying lake of fat-free (and, for that matter, lactose-free) milk. A tall glass of antioxidant-fortified fresh-squeezed orange juice (with one crystal-clear ice cube of distilled water) was also obligatory, as was a large, steaming cup of Jamaican Blue Mountain coffee, black. The kicker was that Suzannah insisted that all this food come from the soil in the old-fashioned way; no replicator-produced victuals for her and her famous husband. Fenwick had long ago given up trying to point out to his better half that molecules were molecules, whether they found their final arrangement over time in an old-millennium garden, or were configured in seconds in a replicator. Suzannah insisted that grown food was healthier, and she pointed out that she’d been right about berries and caffeine. And, she *had* been right: The recommendations that had surfaced at about the start of the new millennium had turned out to have unusual staying power, supported at this point (2029) by a veritable ocean of empirical data. Those recommendations had been to consume cancer-fighting berries, Alzheimer-fighting caffeine, and a host of other things that Fenwick now dutifully ingested every day of his life. (Red wine, also powerfully salubrious in moderation, was something Fenwick was allowed to have only when socializing with other members of the literatí. Fenwick had long ago let fall the objection that his wine cellar, worth a pretty penny, would take a century to enjoy at the Suzannah-prescribed rate of consumption.) When spooning down the cocktail Suzannah prescribed, he routinely fantasized about the day when cures for all forms of cancer would finally be found. On that glorious day, he would be able to revert to a breakfast of cheddar-infused scrambled eggs, sausage, bacon, and greasy home fries, all produced in under three seconds by a replicator.

Suzannah had rather strong prudential reasons for monitoring Fenwick's diet. After all, he was hands-down the greatest living dramatist on the planet, author of 14 screenplays (ten of which had led to films of enormous critical and financial value), nine plays (five of which had galvanized stages in New York, London, Paris, Rome, and Fenwick's hometowns of Oslo and Bergen), eight novels, innumerable poems, countless essays, and, of late, his first biography (on John Quincy Adams, one of the few Americans Fenwick truly admired, and a man with whom, courtesy of the system about to be described, Fenwick had had many face-to-face conversations). But the secret to Fenwick's success, the core of his genius, of course had nothing to do with his diet.

As a man of letters, Fenwick had the Midas touch because he could create characters of utter depth, intricacy, and interest. Once he had the characters created, the rest of his writing, whatever the form, was effortless. But for Fenwick, to "create a character" meant something a good deal more than it meant to the average writer, and in fact it meant much more than what it meant even to other renowned authors of belletristic prose. In short, Fenwick Gibsen would only begin to write the first word when his characters were as real to him as the human beings he knew in the real world. This *modus operandi* had nearly destroyed his marriage.

The problem, very simply, had been that Suzannah Gibsen had been forced to share her home with countless strangers, many of whom were not just somewhat unfriendly, but outright creepy. And the fact that they had been invisible had only made the situation all the more disturbing. Sometimes, in the past, Fenwick would sit beside her, and carry out lengthy conversations with a character in an in-progress work. Of course, only one side of the conversation could be heard by Suzannah: She could hear what her husband said, but she could never hear anything in response. One stormy, ink-black night, with thunder booming and wind howling all around their capacious house, and he and his wife both in bed, lights out, Fenwick had carried out a conversation with someone "standing" in their room. At one point Fenwick had said, with emotion: "All the murders you've committed, and you feel not the slightest remorse?"

Another crack of lightning. The room lit up. But of course no interlocutor was to be seen.

Fenwick again, speaking into the dark: "And what in your background has given you this black heart, my friend?"

Presently, Fenwick said: "Ah. Well, perhaps redemption will be possible. We shall see."

The conversation done, Suzannah now struggled to fall asleep. That was something Fenwick easily managed to do. He lay there in peace, his chest rising and falling smoothly, but she tossed and turned as rain slapped their roof in sheets, and as she tried to throw off the feeling that a third person was in their room: a murderer. The next morning, at breakfast together, Suzannah, her face pallid and weary, whispered: "Fenwick, I can't bear it any longer. Either they go, or I do. With all the money that your writing has brought us, is there no way for you to create them, refine them, interact with them — somewhere else?"

“Suzannah,” Fenwick said, “they live in here.” The author tapped his forehead. “Thus, where I go, they go.”

She looked at him in silence.

Presently she said: “But think about the *words* you write. They live in the external world. You commit content in your mind to computer files, routinely. After all, when you turn over to your editor a play, or a novel, and so on, you turn over text that you have written. I should like to say, sweetheart, that you *externalize* the words that compose your works.”

“True enough.”

“Well,” she continued, “then either you find a way to externalize your characters, and circumscribe your interaction with them, just as you do with the text you write, or I go.”

The next day, Fenwick, calling in favors from some rather powerful friends, conversed by phone with one of the greatest neuroscientists in the world, at Columbia-Presbyterian: Francis Sullivan.

“Fenwick, no,” Sullivan had answered. “I’m sorry. We cannot extract what you imagine when you form characters for a story. We could know inside and out what regions of your brain are active, down to the molecular level. But how, pray tell, would you have us extract emotions and personality from patterns of neural circuitry flowing through time? You can’t derive semantic content from mere syntax. This is now an axiom in neuroscience. For example, we can tell when a person is lying by scanning his brain while we ask him key questions, but we can’t tell what the *content* of the lie is. And certainly we shall never be able to see what it a person is *feeling* when lying.”

Fenwick simply said, “I see,” in resignation.

“You’re talking to the wrong kind of expert in me, Fenwick. Talk to someone who can tell you how to speak your characters into existence before your eyes; and someone who can tell you how to interact with them after they have been born.”

“What field would such a person be in, Dr. Sullivan?”

“Your characters are intelligent creatures, right?”

“Certainly,” Fenwick replied.

“And they are artificial, correct?”

“Sure.”

“Then it seems quite plain: Go consult the AI folks.”

Fenwick had done so. And now, after an investment of $40 million to build a home office more sophisticated (but more intimate) than the only three holodeck-like installations he had found (in laboratories at RPI, MIT, and Georgia Tech), all his marital turbulence was in the past. Thanks to advances in AI (of a particular variety; we’ll get to that), and also to holography, connubial bliss now reigned in the Gibsen home, and the master writer was more productive than ever.

After breakfast on this particular morning, Fenwick pulled on his fur-lined boots, put on and buttoned up his sheepskin coat, pulled his wool cap down to cover his ears, and slid on his mittens. He then departed, and trudged off into the snowy woods, bound for his office, a stone cottage on his own property set back one mile into the forest. From the outside, from the fieldstone walls and slate roof, and the stone chimney from which curled a plume of smoke from the fire burning inside, the cottage looked to be anything but high-tech. Even inside, the only tip-off to the peerless technology that Fenwick had bought was to be found in those with whom the great writer interacted.

The second Fenwick stepped inside his cottage, Arnie greeted him: "Hello, Fenwick."

"Arnold, hi. Listen, I've been thinking. It seems to me that you're appearance is too conventionally, well, too conventionally *threatening*. It's true that you're a rather unsavory interrogator in the story, but I don't want your physical appearance to be a stereotype."

"What do you suggest?" Arnie asked.

"Well, for starters, you should be shorter, much shorter."

Immediately Arnie shrunk a full foot.

"And thinner, bordering on frail," Fenwick said.

Arnie lost 25 pounds in an instant.

"With less hair," Fenwick directed. "A *lot* less."

Arnie's hair thinned out to the point that his skull could be clearly seen through what was left. The process continued until baldness arrived.

"And I think your physiognomy, while not engaging, should be a bit more … *pleasant*."

Arnie's face began to morph. His brow receded. But the same intelligent forehead remained, as did the prominent hook nose.

"What do think?" Arnie asked.

Figure 1: Arnie, Face View

Fenwick studied Arnie for a minute, walking slowly around him in a circle. “Yes, I like it,” Fenwick said.

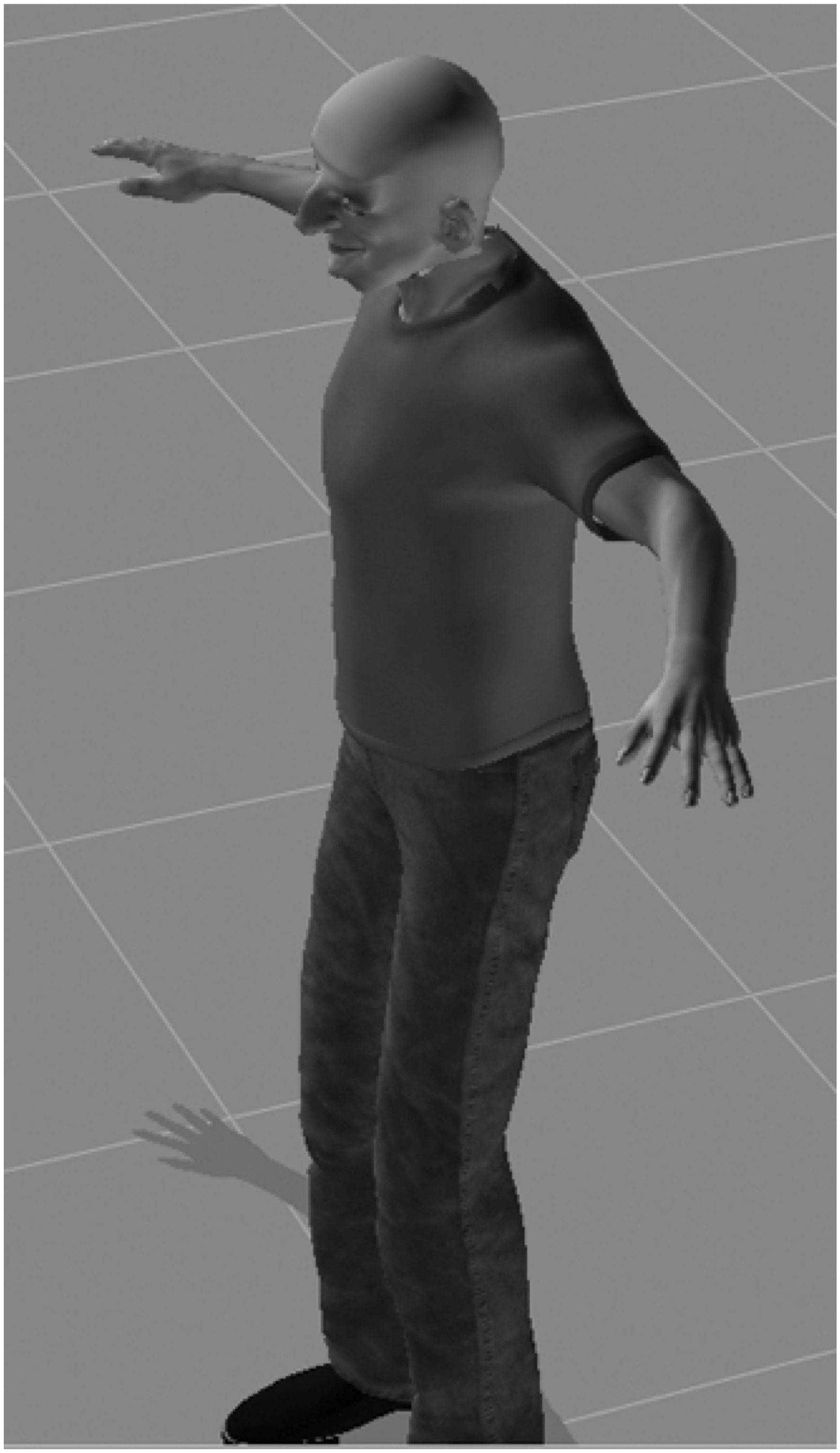

Figure 2: Arnie, Full-Body

And then he continued: “Your voice is too low. I think you need a soft voice, one that borders on a whisper, but suggests, subtly, a disquieting hiss.”

Arnie spoke: “Very well, Mr. Gibsen. How do I sound now?”

"Perfect," Fenwick replied.

At present, Fenwick was working on a play, *Calculi of Death*, or as Fenwick referred to it: *Calculi*. The title reflected the fact that in this play, a number of people run rather different calculations in order to try to figure out how fierce interrogation of a certain prisoner can become, before the interrogation becomes unethical. The prisoner, Muhammad, knows the secret location of ticking nuclear bombs planted within the mainland US. If these bombs are not diffused, millions of civilians will perish. If Muhammad discloses the locations before a certain deadline, US authorities can diffuse the devices, guaranteed.

Interrogators in *Calculi* include General Farrow and Arnie, with advice—ethical and theological — offered by the Reverend Franklin Gray. The President of the United States, President Bentham, is also one of the characters, and has ultimate say as to whether or not Arnie is allowed to proceed as he wants. General Farrow's view of interrogation, reflective of the standpoint of Occidental culture, is fairly standard: He refuses to use unadulterated torture, even when the stakes are as high as they are in this case; and even "tough tactics," such as a punch to the face, while for him admissible, are things he hates to even ponder, let alone implement. Gray is of the opinion that the Judeo-Christian ethic that he has devoted his life to upholding clearly implies that torture is unethical—and points out that if one uses a calculus that permits unlimited pressure to be applied to Muhammad, both Judaism and Christianity are vitiated, since the message being sent by such violence is that God isn't in control after all, and that therefore matters must be managed instead by human beings.

Arnie is quite a different animal. If he sees *anything* in the behavior of a prisoner holding such life-and-death secrets that indicates the prisoner will yield to certain extreme techniques for extracting information (and we cannot share what these techniques are here; readers are directed to the following url for the latest draft of *Calculi*: http://kryten.mm.rpi.edu/Calculi_of_Death.pdf), however painful those techniques may be, he will deploy these techniques without the slightest qualm. Below, we will look at what is happening from a technical point of view when Arnie sees a behavioral cue that triggers his pouncing, full-force, on Muhammad.

Discussion

Creative writing can be viewed from the perspective of the degree of content that can be *externalized* by the human author outside the mind/brain of that human author, courtesy of the best technology of the day. (Confessedly, this is a view of creative writing reflective of the fact that our interests, in the end, are more in the engineering direction, rather than the literary one.) Go back a bit, and authors didn't have computers, so they didn't have electronic text processing systems that are in the very bloodstream of today's writers. Creative writers, however, always had *something* to write on. It may not have been *easy* for the first creative writers to write text on an external object, and then rework that text—but it was done. On the other hand, there has always been a natural evolution of increasingly more sophisticated ways of externalizing text, and of allowing that text to be modified, refined, and so on. In addition, creative writing has always had both pictorial and auditory elements, and in light of the fact that in the history of hominids, there is evidence of the expression of narrative before language, we are really talking about the *tri*-evolution of technology for externalizing linguistic, auditory, and pictorial content related to the narrative in question. The state-of-the-art today in this tri-evolution is perhaps that which is available to screenwriters and playwrights (and also to novelists, but they tend not to use the software in question, for some reason), in the form of software systems that allow for the representation and manipulation of text, meta-representations of that text (outlines, storyboards, etc.), and visual and auditory tools.[1] That which is available to Fenwick Gibsen is really only a natural extrapolation along the evolutionary arc that creative writers have been travelling steadily upon for many, many centuries. Since Gibsen, like his namesake the brilliant Henrik Ibsen, is a dramatist who writes from a core composed of rich, robust characters, rather than from plot, technology for externalization in his case is devoted to enabling the construction and refinement of well-rounded characters. As Ibsen famously said, he would not take up pen for any play until he had all his characters in mind down to the buttons on their clothes (Bringsjord 1995).

We turn now to the theoretical foundation for the synthetic character/world technology exploited by Fenwick Gibsen. Much, in fact arguably most, of this foundation has already been erected as of 2009. We will pay special attention to the two cornerstones of this foundation that will likely prove to be the hardest to fully perfect: the part supporting real-time natural-language communication between a creative writer and his/her characters, and the part supporting the high-level cognition of those characters, specifically their beliefs about the minds of others. We begin by discussing the fact that both of these cornerstones of the foundation pertain not to the physical appearance of characters, but rather to their cognitive powers.

Theoretical Foundation For This Future

Clearly, given the goal of building an environment of the sort that Fenwick exploits in order to write creatively, there would appear to be little need for us to look carefully at the *biological* basis of human persons. The main reason is that the physical side of the future we imagine, even now, is remarkably close to what is currently available. The physical design of Arnie, as shown above in Figures 1 and 2, is already something that can be to a high degree carried out; indeed, we *have* carried it out. Arnie's physical

---

[1] Our current favorite such system is *Movie Magic Screenwriter*, which (e.g.) allows the writers to hear their characters speak to them in desired styles. A succinct overview of the dominant systems in this class is provided in "Product Review" by Sean Kennelly, on page 74 of *Creative Screenwriting* (September/October 2008). His review specifically targets *Movie Outline v3*, which we also have, and very much like. *MO3* has built-in functions for building characters, and can be considered a primitive precursor to Gibsen's ability to modify Arnie (and other characters) merely by speaking commands.

appearance has been mapped out using the software system *Poser 7*—the '7' deriving from the fact that this is the seventh version of the system.[2] Now of course Arnie, as of 2009, appears to the authors of the present paper in 2D form. (Dropped into a game engine, Arnie does appear as a 3D character, but that is of course a "fake" 3D view. Fenwick, on the other hand, quite literally sees Arnie the same way he sees his wife: in full-blown, real-world 3D.) For interaction of the sort described above between Fenwick and Arnie, a full, life-like, 3D representation of Arnie is necessary, and hence it can be safely said that advances in holography and related areas (e.g., projection onto 3D shapes) will be needed if the future we have described above is in fact to arrive. We herein ignore this dimension of Arnie. Frankly, it is a dimension that is infinitely easier to advance, engineering-wise, than the cognitive and linguistic dimensions.

What's important to Fenwick about Arnie, and other characters in *Calculi of Death*, is their *cognitive* side. Even in cases where a character's physical status is central (e.g., in situations where a character is seriously ill, and the illness plays a significant role in the drama), it is the mental space that is of paramount importance. In short, when we look to the theoretical foundation of the system Fenwick uses, we must expect that foundation to relate to computational psychology, not to, say, the DNA structure of humans. When it is said that *x* is human just in case *x* has a particular genetic code, the perspective is not that of computational psychology. Along the same lines, our minds are not modeled by charting the physiology of our brains. Rather, the builders of Fenwick's system must be concerned with what it means to be a human being, from the *psychological*, and indeed specifically the *cognitive*, perspective. That is, the question is: What does it mean to be a human *person*? For ambitious AI, the centrality of personhood is plain in the relevant literature. For example, here is the more than two-decade-old objective for AI (Charniak and McDermott 1985):

> The ultimate goal of AI, which we are very far from achieving, is to build a person, or, more humbly, an animal. (Charniak & McDermott 1985, p.7)

More recently, other authors have made plain that AI aimed explicitly building persons is very much about the incorporeal, or purely intellectual side (e.g., see Pollock 1995).

One generic account of human personhood has been proposed, defended, and employed by Bringsjord (1997). This account is an orthodox one; for example, it generally coincides with one given by Dennett (1978), and by others as well, for example Chisholm (1976). In addition, this account is in line with the capacities covered, chapter by chapter and topic by topic, in surveys of cognitive psychology (e.g., see Goldstein 2005). The account in question holds that *x* is a person provided that *x* has the *capacity*

1. to "will," to make choices and decisions, set plans and projects — autonomously;
2. for subjective consciousness: for experiencing pain and sorrow and happiness, and a thousand other emotions — love, passion, gratitude, and so on;

[2] Any number of other options are available today. *3D Studio Max* is another nice tool, and various games/game engines offer tools. E.g., we have *Half Life*'s character design tool running in our laboratory.

3. for *self*-consciousness, for being aware of his/her states of mind, inclinations, preferences, etc., and for grasping the concept of him/herself;
4. to communicate through a language;
5. to know things and believe things, and to believe things about what others believe (second-order beliefs), and to believe things about what others believe about one's beliefs (third-order beliefs), and so on;
6. to desire not only particular objects and events, but also changes in his or her character;
7. to reason (for example, in the fashion exhibited in the writing and reading/studying of this very chapter).

Arnie (and, for that matter, other characters in Fenwick's *Calculi of Death*) appear and interact with Fenwick on the strength of work carried out to capture parts of this definition in computation. Notice that we say *parts* of this definition. The fact of the matter is that some of the capacities constitutive of personhood don't seem to be amenable to mechanization, even in the long term. This seems to hold for Attribute 2 (Yang & Bringsjord 2003, Bringsjord 1999, Bringsjord 1998), and we discuss this issue at the very end of our chapter.

Anderson and Lebiere (2003), instead of defining personhood, give an operational equivalent of this definition by describing what they call "Newell's Program," an attempt to build computational simulations of human-level intelligence, where that intelligence is cashed out in the form of a list of abilities that correspond to those on the list just given. For example, part of Newell's Program is to build a computational simulation of natural-language communication at the normal, adult level. This is Attribute 4 on the list above. As Anderson and Lebiere concede, AI and cognitive science (CogSci) are finding it rather difficult to mechanically simulate this attribute.

In the present chapter our emphasis is on Attributes 4 and 5, as we have already indicated. That is, we are concerned to explain how it is that the theoretical foundation for Arnie's having the capacity to communicate in English, and his having the capacity to know and belief things (up to at least third-order beliefs) on the strength of what he perceives, is in large measure already in place. (We also seek to prescribe a path forward to build atop this foundation. More on that in due course.) We turn now to the two cornerstones in this foundation.

Cornerstone #1: Natural Language Communication/Conversation

Twenty-five years ago, the dream of engineering an artificial agent with the capacity to communicate in a natural language (e.g., English, German, Norwegian, Chinese) at the level of human persons was alive and well. After all, Alan Turing, who can plausibly be regarded the founder of modern-day AI, inaugurated that field with a landmark paper in which he identified thinking with linguistic performance, and also predicted that before the year 2000, computing machines would be linguistically indistinguishable from neurobiologically normal humans (under certain reasonable time constraints; see Turing 1950). In addition, when AI officially started in the US in 1956 (at a DARPA-sponsored workshop at Dartmouth), Newell and Simon took themselves to be ushering in a research-and-development path that would very rapidly produce programs with human-level eloquence (see the quotes in Russell & Norvig 2002). As far as we can tell, the dream of human-machine linguistic equivalence was not based on any particular piece of technology, nor on any particular theoretical or mathematical advance, but rather simply on the assumption that natural language could be converted, without loss of meaning, into corresponding content expressed in a formal language. For example, as described in (Charniak & McDermott 1985), sentences in English were to be converted into formulas in the language of first-order logic (FOL). The basic assumption that communication in natural languages could be translated without loss of meaning into a formal logic-like language can be traced back much earlier than 1985: it can be traced back to Leibniz, who asserted in 1685 that disputes would eventually be settled by calculation in a formal language (Weiner 1951).

We haven't the space to provide a full review of FOL, or for that matter of any other logical system.[3] Readers wanting such a review can turn elsewhere (e.g., to Bringsjord & Yang 2003; and for an extensive presentation of logical systems in general, including FOL, all given in the context of precisely the kind of "mind modeling" that bringing Arnie to digital life requires, to Bringsjord 2008b). Suffice it to say here that FOL is based upon the simple idea that objects can have properties, and that these properties can be binary, ternary, quarternary, and so on. To say that Selmer, denoted by the constant `s`, has the binary property of being older in age than Alexander, denoted by the constant `a`, we can write in FOL:

```
Older(s,a).
```

In this simple formula, `Older` is a predicate. Functions can also be written in FOL in straightforward fashion. For example, let's assume that `age` is a function that returns the age of a constant (when that constant names a person), so that we can write such things as `age(a)`, which (at present) returns 19. In addition, FOL has the Boolean connectives that all readers will at some point have seen during their education: $\wedge$ ("and"), $\vee$ ("or"), $\neg$ ("not"), $\rightarrow$ ("if … then …"), and $\Leftrightarrow$ ("… if and only if …"). FOL also has two *quantifiers*: the *existential quantifier* $\exists$, which, when paired with a variable `x`, as in $\exists$`x`, says "there exists at least one thing `x` such that …"; and the universal quantifier $\forall$, which, when paired with a variable `x`, as in $\forall$`x`, says "for every thing `x` …" Given this machinery, how would we represent in FOL the following sentence?

---

[3] Note that there is a profound difference between a *formal language*, and a full *logic*, or *logical system*. The former requires only an alphabet of primitive characters, and a grammar for assembling well-formed or grammatically correct formulas from that alphabet. The latter, in short, requires not only such *syntactic* machinery, but also a formal *semantics* that specifies under what conditions the formulas have such values as *true*, *false*, *indeterminate*, *probable*, *certain*, and so on. In addition, a logical system includes some framework for systematically reasoning over the formulas; such reasoning can come in any of a number of modes, which include deduction, induction, abduction, and analogical inference. For an overview of logical systems, see (Bringsjord 2008b). In Bringsjord's account, a logical system also includes meta-theorems that provide key mathematical information, such as whether or not all formulas that are true no matter what they represent can also be proved.

The age of every father is greater than than that of all of his sons.

As follows:

```
∀x∀y(Father-Of(x,y) → Greater(age(x),age(y)))
```

And for an example related directly to *Calculi*, observe that the sentence

If Arnie deploys his interrogation technology, Muhammad will divulge his secret.

could in FOL be represented by something like the following first-order formula:

```
Deploys(arnie,i-tech(arnie)) → Divulge(muhammad,m-secret)
```

And the more general English sentence

Anyone who deploys Arnie's interrogation technology will get Muhammad to talk.

becomes

```
∀x (Deploys(x,i-tech(arnie)) → Divulge(muhammad,m-secret))
```

As it happens, our logic-based AI work, and indeed the vast majority of such work carried out in our laboratory, is based upon an extremely useful variant of FOL: *multi-sorted logic* (MSL). A comprehensive presentation of MSL can be found in (Manzano 1996), and an example of R&D in our lab that leverages MSL can be found in (Bringsjord, Taylor et al. 2008). The basic idea behind MSL is that the set of objects described via formulas is partitioned into sub-categories. For example, one might have a sub-category (or *sort*) composed of people, and another composed of numbers, and so on.

Why would anyone want to reduce natural language to formulas in FOL or MSL? The answer is that from the dawn of AI, certainly one of the most impressive powers possessed by computing machines is their ability to reason automatically over propositional content expressed as formulas in logical systems.[4] In the painfully simple example of Arnie and Muhammad, a machine overseeing a knowledge base that includes the proposition that Arnie does in fact deploy his interrogation technology, along with the two formulas shown immediately above, can of course instantly deduce that Muhammad will divulge his secret. In general, if we assume that the collection of all that Arnie knows and believes (and fears, and hopes, etc.; these verbs all take as targets propositions and are known as *propositional attitudes*) can be identified with some set Φ of formulas in MSL, and we assume that any natural-language question uttered by Gibsen is expressed in MSL by some query φ?, then part of the cornerstone we are currently describing consists in the fact that machines can often answer such queries by attempting

[4] FOL is one logical system. The space of all such systems is infinite, and includes many different logics. For an explanation of the concept of a logical system, see (Bringsjord 2008b).

to find a proof of φ from Φ. One of us has explained this basic scheme in much detail elsewhere (Bringjord 2008b).

To this point, despite the richness of logical systems, natural language has resisted attempts to model it in logico-computational terms. However, the attempt to model *proper subsets* of unrestricted natural language in formal logic (e.g., see Fuchs 1999 and Bringsjord et al. 2007) has been quite successful, and it is this research activity that forms Cornerstone #1 of the theoretical foundation for Gibsen's remarkable home office. Specifically, the key research activity, as outlined in Figure 3, is the three-stage process used in our laboratory by which propositional content expressed in unrestricted natural language (UNL) is eventually expressed in MSL. In the first stage of the process, UNL is translated into LCNL. In the second stage, LCNL is translated into corresponding *discourse representation structures*.[5] In the third and final stage, these structures are translated into formulas in MSL. In our lab, the second and third stages are mature and reliable, automated, and quite well understood. The first stage, on the other hand, is acutely challenging, and is in need of considerable improvement, in light of the fact that this stage currently only works when input sentences in UNL are "well-behaved." We will discuss the first stage this below, in the section "Prescription for Future R&D."

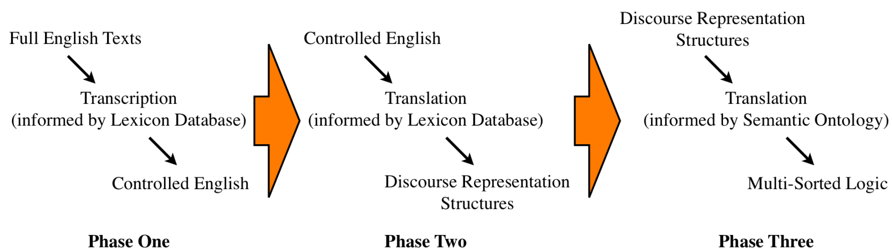

Figure 3: Overview of Logic-Based Machine Reading Process

Cornerstone #2: A Theory of the Minds of Other Minds

The principles and techniques that neurobiologically normal humans deploy in order to understand, predict, and manipulate the behavior of other humans are collectively referred to as a *theory of mind* (Premack & Woodruff 1978). In *Calculi*, the fact that Arnie has a theory of mind (ToM) is what underlies his beliefs about Muhammad's beliefs, and indeed his beliefs about what Muhammad believes about what Arnie himself believes. Before looking in some detail at an example of ToM-supported reasoning on the part of Arnie, we briefly discuss an example that has been explored in our laboratory. This example has been based on the much-discussed *false-belief test* (Wimmer & Perner 1983).

In the false-belief test, a subject *S* faces two containers, A and B. The experimenter, with *S* looking on, places an object *o* into A. Not only does *S* see that *o* is

---

[5] DRSs are a syntactic variant of first-order logic for the resolution of unbounded anaphora. Their use in the interpretation of text is a central element of *discourse representation theory* (Kamp & Reyle 1993; Kamp & Reyle 1996), upon which we directly rely. Discussion of this theory is beyond the scope of the present chapter.

placed into A; he also sees that another person, *S'*, sees that *o* is placed into *A*. At this point in the test, *S'* is asked to leave. While he is away, *S'* cannot see what is taking place in connection with the containers and the object, and this limitation can be observed by *S*. After *S'* leaves, the experimenter, again in plain view of *S*, moves object *o* from A to B. At this point, *S'* returns, and the experimenter asks *S*: "Now *S*, if I ask *S'* to get *o*, which box will he look in, A or B?" A subject passes the false-belief test if he or she correctly answers that *S'* will look in container A.[6] The subject fails if he or she says that *S'* will look in B. Almost all children before the age of four fail the test. Children above this age almost invariably pass.

In Figure 4, the center synthetic character, Eddie, who is playing the role of *S*, is shown, and to his right stands Micah, playing the role of *S'*. In the snapshot shown in Figure 4, the question shown is asked by the experimenter, who stands to Eddie's left. Of course, the action that took place before Eddie asked this question included the initial dropping of the teddy bear into container A and the transfer of the bear from A to B while Micah was away. Figure 5, in its insert, shows a snapshot of the automated reasoning that corresponds to the thinking taking place in Eddie's mind. This reasoning takes place on computers in our laboratory, despite the fact that the relevant animation happens in *Second Life*. This reasoning takes place in a "socio-cognitive" calculus more expressive than FOL and MSL. The increased expressivity is required in order to allow for formulas corresponding to such propositions as

Eddie believes that Micah believes that the teddy bear is in Container A.

If we were to try to directly represent such a sentence in FOL, we would get something like this:

```
Believes(e,Believes(m,In(t-b22,container-A))).
```

But this is flat-out grammatically incorrect in FOL, because predicates cannot have predicates, formulas, or sub-formulas as arguments in FOL. Indeed, no such constructions are permitted in any logics that are (as it is said) *extensional* in nature. Traditionally, extensional logics are developed for the purpose of formalizing domains quite separate from cognition—domains like number theory, which posit only objects bereft of psychological attitudes like believing, knowing, intending, and so on. At the opposite end of the spectrum comes creative writing, which is nothing without highly elaborate cognition on the part of characters in narrative, readers of that narrative, and writers of that narrative. This is something one of us has explained in connection with murder mysteries; see (Bringsjord & Ferrucci 2000). In murder mysteries, typically a clever detective hot on the trail of a villain who knows he is being pursued will have

---

[6] In our version of the test, a subject passes only if he or she not only answers correctly, but can also supply the correct justification for this answer. A subject who simply answers A or B based upon, say, the flipping of a coin, would not pass the test, and for us, this fact is important. Unfortunately, in the cognitive psychology literature, subjects are rarely asked to provide an argument in support of their answers.

beliefs about the beliefs the villain has about the beliefs of the detective. For example, it may well be that the detective believes that the villain believes that the detective believes the villain has such-and-such a property. And given this, it's easy to see that the villain could discover it, and hence it would be the case that the villain believes that the detective believes the villain has such-and-such a property. For the details regarding how these kinds of constructions are representable, and how they can be rigorously reasoned over in the socio-cognitive calculus, see (Arkoudas & Bringsjord 2008).

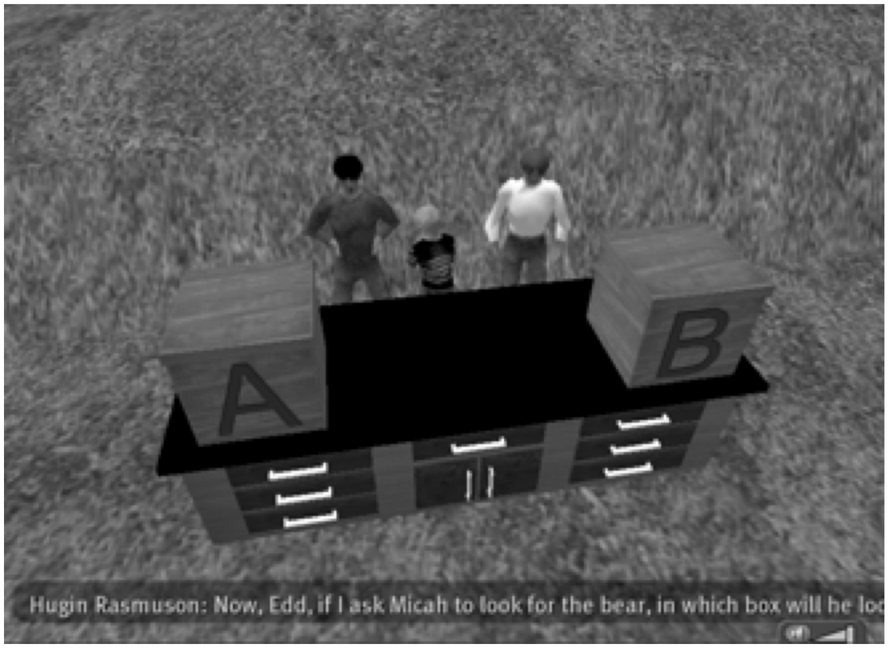

Figure 4: Eddie Just Before His ToM-Based Reasoning

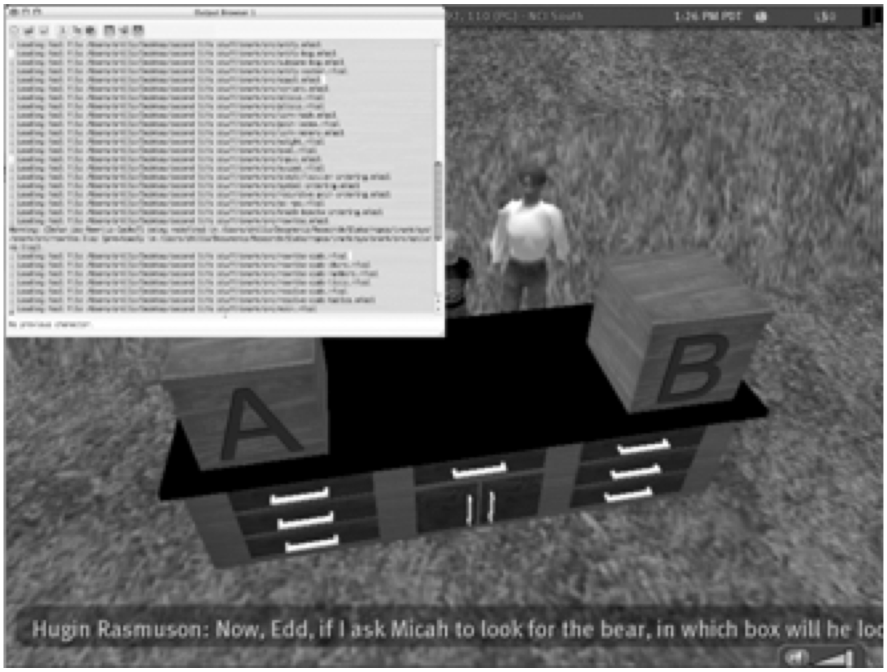

Figure 5: Eddie During His ToM Reasoning

But these examples don't pertain to the case of hand, which involves Gibsen as the author of *Calculi*, and its most complex character: Arnie. We now provide an example of Cornerstone #2 in action, in connection with Arnie. To begin, we shall assume that it is well-known that certain behaviors indicate that those performing them are at least somewhat fearful. There are many candidate behaviors—cringing, letting one's eyes go wide, shivering, shuddering, and so on. To fix things in our example, for no particular reason we shall pick the action of taking a deep, cleansing breath while at the same time squinting. This action will be taken while one is under interrogation. We shall assume that this is a tell-tale sign that the person performing this action is at least somewhat fearful, and that it is common knowledge that if an agent believes that he has performed this action, that agent believes that he is indeed fearful. We shall denote the particular composite action in question by `dbreath`. We shall further assume that there is some formula $P^{fear}(a)$ in FOL that says that some agent `a` is fearful. Note that this might be itself a rather complicated formula, so we are denoting it with an abbreviation.

We can now generalize a bit: We can let α^* be a variable that can have as an instance any particular fear-indicating action, not just `dbreath`; and we can let $P^*(a)$ be a variable that can have as an instance any psychological state, not just fear. We are generalizing in this case to set up a framework in which we can formalize and mechanize the fact that human beings, upon seeing any of a vast number of behaviors, instantly infer that the bearer of this behavior is in some particular mental state. The generic procedures that produce such chains of inference are called *methods* (Arkoudas & Bringsjord 2008).

Please refer to Figure 6 to see the actual chain of inference that underlies part of what Gibsen has written: a scene in which Arnie, upon seeing Muhammad take a deep breath and squint, believes that Muhammad is fearful, and indeed believes that he

Figure 6: <INSERT PROPER CAPTION>

Sample Method-Generated Proof in "Cognitive Calculus" of One of Arnie's Beliefs

Set the particular action-type α^* to *dbreath*, and the time of the dropping to t^*. In addition, let $P^{fear}(a)$ stand for the formula expressing the proposition that agent a, the only free variable in P^{fear}, is fearful.

The method (let us call it *M*) in this case allows a psychological state of the agent a to be quickly deduced from the performance of the action α^*. With a_a representing Arnie, and a_m representing Muhammad, what happens when the method fires is that the following proof is instantly generated. Note that the conclusion of this proof is that Arnie believes that Muhammad believes that Muhammad himself is fearful. Given this belief, Arnie's goal is closer, and he is emboldened.

Proof:

1	$\mathbf{S}(a_a, happens(action(a_m,\alpha^*),t^*))$	premise
2	$\mathbf{K}(a_a, happens(action(a_m,\alpha^*),t^*))$	1, DR_4
3	$\mathbf{K}(a_a, \forall a,\alpha,t.happens(action(a,\alpha),t) \Rightarrow \mathbf{K}(a, happens(action(a,\alpha),t)))$	A_4, DR_2
4	$\mathbf{K}(a_a, happens(action(a_m,\alpha^*),t^*) \Rightarrow \mathbf{K}(a_m, happens(action(a_m,\alpha^*),t^*)))$	3, DR_9
5	$\mathbf{K}(a_a, \mathbf{K}(a_m, happens(action(a_m,\alpha^*),t^*)))$	4, 2, DR_6
6	$\mathbf{B}(a_a, \mathbf{B}(a_m, happens(action(a_m,\alpha^*),t^*)))$	5, DR_5, DR_{12}
7	$\mathbf{C}(\forall a,t.\mathbf{B}(a, happens(action(a,\alpha^*),t) \Rightarrow P^*(a)))$	premise
8	$\mathbf{K}(a_a, \forall a,t.\mathbf{B}(a, happens(action(a,\alpha^*),t) \Rightarrow P^*(a)))$	7, DR_2

believes he is. (Muhammad actually *knows* that he is fearful, but we leave this aside.) In this proof, provided for readers passionately interested in the details of our approach, the operator **B** stands for *believes*, **K** for *knows*, **C** for something that is common knowledge (= known by all agents), and **S** indicates *sees*. There are only two premises in the proof; they are labeled as such in the rightmost column. Every other line is machine-deducible. We don't have the space to explain each inference, and explain only the one made at line 2: Line 2 can be deduced using the general rule of inference (DR_4) that essentially says that whatever an agent sees an agent knows. So in this case, since Arnie sees Muhammad's tell-tale deep breath and squinting, he knows that Muhammad has indeed taken such a breath.

Prescription for Future R&D

Can we really get there from here? Is it really true that today's logic-based brand of AI and CogSci has in fact laid the foundation for providing creative writers with a work environment like that which Gibsen enjoys? If so, how do we get from the groundwork, from the cornerstones described above, to the wondrous home office in question? There are three sustained projects that must be launched in order to forge a link from today's foundation to Gibsen's holodeck-infused home office. As will be seen, the first two of these projects both require a sustained investment in the refinement of LCNL-based work, while the third involves applying parallelization techniques that are now common in non-cognitive supercomputing domains to the mechanization of ToM. We now briefly describe these three projects in turn.

Project #1: Boost LCNL to Cover Doxastic and Other Operators

As we explained above, logical systems equipped to represent doxastic operators (e.g., *believes* and *knows*) already exist, and in some cases are accompanied by precise proof theories that make automated reasoning over such propositional content possible. Such is exactly the case with respect to the socio-cognitive calculus of Arkoudas & Bringsjord (2008). LCNL, however, is currently only able to capture the structure and meaning of sentences devoid of such operators. In order to reach the day when Arnie will be not just a character in a vignette like the one we opened with, but a cognitively robust synthetic character able to interact with real creative writers, it is essential that LCNL be expanded to cover these operators—as well as other aspects of unrestricted natural language that currently cannot be formalized and mechanized in logic-based AI or CogSci.

What is involved in the push to augment the expressivity of LCNL so as to cover doxastic operators? Developing a detailed answer to this question would constitute an extensive research program in its own right. Nonetheless, we can venture a few general points in response to the question.

The first point is that the traditional semantics for logical systems created to represent doxastic operators seems to us to be ill-suited to providing guidance for how to augment and modify LCNL to handle such operators. Along with Arkoudas, we are of the view that the meaning of formulas involving the **B** and **K** operators from above should not be given in terms of traditional so-called "possible worlds," which is common in philosophy and AI (as for example in Fagin et al. 2004), but rather in terms of the rationales that support the relevant attitudes. In other words, on this approach what it means to say that some agent believes a proposition (at least when this belief is a conscious one) is that there is a web of inference conferring plausibility on this proposition, from the standpoint of what the agent knows. So, you believe that Lincoln was President of the US, and the meaning of this belief consists in the fact that there are many things you know that can be arranged essentially into an elaborate series of (not necessarily deductive) arguments for the proposition that Lincoln was President. (For more on this general direction, see (Bringsjord, Shilliday et al. 2008).) The advantage of this approach is that it should be much easier to associate concrete algorithms with belief, since if there is any success story within AI it surely must be the amazing progress machines have made in processing proofs and arguments.

Project #2: Capture an Appreciable Percentage of the Meaning of UNL in LCNL

Suppose for the sake of argument that Project #1 is pursued aggressively, and is a resounding success. This means that the range of LCNL will have been significantly extended. But there is still a severe challenge: viz., it is not possible to mechanically translate arbitrary UNL into LCNL in such a way that preserves meaning. In fact, the situation is quite dire, in light of the three-phase process discussed above, and shown in Figure 3. As the reader will recall, the second and third phases are already automatic. This immediately implies that if there was on hand a way to translate UNL to LCNL the entire natural language problem in AI would be solved! But that problem is enormously challenging (a fact we return to in our conclusion, below), and we believe that creative writers can be empowered by synthetic character/world technology long before this problem is solved (if it ever *can* be solved).

The solution is to aim at capturing only a *percentage* of the content of text expressed in UNL when that text is translated into LCNL. To our knowledge this approach has never been explored. If Gibsen suddenly began to engage in conversation that breaks the bounds of LCNL, the technology underlying Arnie should take some risks in recasting Gibsen's language into LCNL. Consider tense, for example. Currently, LCNL doesn't provide a way to represent sentences using anything other than the present tense. But as one of us can tell you, even when you can only speak a foreign language in the present tense, you can still carry out rather elaborate, productive conversations with native speakers of that language.[7] We do not ourselves yet have on hand algorithms for moving from UNL to LCNL in a manner resigned to losing considerable information in the process. What we are saying is that the development of such algorithms, and experimentation therewith, is a project that is part and parcel of our roadmap for how to get from today, to what Gibsen enjoys in 2029.

Project #3: Use Methods to Parallelize Automated Reasoning over Content in Socio-Cognitive Calculus

The third and final project we recommend is the parallelization of the machine reasoning central to our approach. Please refer back to Figure 6, and the proof shown therein. As we have already explained, the fact that this proof is produced quickly by a so-called

---

[7] Selmer never managed to move beyond the present tense in speaking and writing Spanish, but found ways of nonetheless communicating with Spaniards.

method, and is not laboriously constructed *de novo* for the particular case at hand, is crucial. The idea that human persons go around consciously constructing, step by step, complicated chains of reasoning like what is shown in Figure 6 is not psychologically plausible in the least. That is why methods are so attractive, for on the other hand that human persons have on hand *general-purpose procedures* (which are essentially what methods are) for rapidly forming inference-supported beliefs isn't psychologically implausible in the least. But with methods on the table, the next natural step is to allow agents to run methods *in parallel*. At the same time that Arnie is looking for, and making inferences from, the facial expressions and gestures of Muhammad, he must be attuned to the attitudes of his superiors (who are monitoring his behavior), to words Muhammad utters, to words he himself utters (and to the strategies that dictate what he ought to say), to his own premeditated facial expressions and gestures, and so on *ad indefinitum*. All of this, and all of the reasoning that accompanies it, must be carried out in parallel. That seems obvious, and yet automated theorem proving has been steadfastly sequential, not parallel. The reason for this is that the basic formats for automated reasoning, for example reasoning based on the rule of inference known as *resolution*, are not ones that are easily parallelized.[8] The situation is quite different in the case of methods: There would appear to be no major obstacle standing in the way of a system capable of simultaneously firing multiple methods. To our knowledge, this is a research direction that has yet to be explored, but it is one that is crucial to our vision.

Two Caveats

Finally, what pair of caveats have we in mind? Well, these:

First, however painful it may be, and despite what we have said above, we need to be dead honest about the unmitigated difficulty of building artificial agents able to deal with *unrestricted* natural language in ways that are indistinguishable from those used by normal human persons of average intelligence. While as we have noted, Turing (1950) predicted over half a century back that by now we would be able to engineer machines linguistically indistinguishable from us (i.e., machines able to pass his so-called "Turing Test"), the fact of the matter is that, today, a bright toddler's conversational reach still exceeds that of any and all computers on our planet. In the approach described above, there is a loss, and indeed a *significant* loss, that takes place when moving from completely unrestricted natural language to logically controlled natural language. We believe that if this loss is small enough, natural-language communication between a creative writer and his or her characters can be sufficiently robust to enable that writer to work more efficiently, and perhaps at a higher level of quality. But we are under no illusions about how hard it is for a machine to communicate at the level of a human person.

[8] Nice coverage of resolution-based theorem proving can be found—within the context of AI—in (Russell and Norvig 2002).

Of course, there are those (e.g., Moravec 1999) who hold that, relatively soon, person-level communication will be mechanized. Unfortunately, such writers are confident because of the continuous increase in processing speed produced by Moore's Law, but raw processing speed is not the problem (as explained in Bringsjord 2000): the challenge is to discover the information-processing *procedures* that enable human persons to communicate in unrestricted natural languages. However fast the hardware, it does little good unless there are procedures to run upon it. It can therefore be said with confidence that while logically controlled natural language, and clever translation of unrestricted natural language into it, will pay great dividends, we certainly don't recommend that you hold your breath in anticipation of the arrival of human-to-synthetic-character linguistic interaction indistinguishable from the human-to-human case. Fenwick can talk to Arnie within the confines of the narrative structure that Fenwick has himself established. He can't talk to Arnie about anything, in any manner. It's ironic that some originally logic-based experts in computational language processing have turned their backs on logic, in favor of purely statistical approaches. Charniak himself is an example; he abandoned logic in favor of purely statistical approaches (Charniak 1993). The situation is ironic because statistical approaches are content to capture only a portion the natural language such approaches receive as input, and yet in the present chapter we are recommending a logicist path based on the humble concession that only a portion of natural language received by Arnie as input can be transformed into logically controlled natural language. Of course, the nice thing about the approach we recommend is that that which is expressed in logically controlled natural language can be reasoned over with great power.

And what of the second caveat? It is that we need to be crystal-clear about the apparent insurmountability of qualia; that is, about the outright impossibility of literally imbuing synthetic characters with the capacity to feel what it's like to, say, drink a great Pinot Noir, or carve a ski turn on hardpack at 40 mph, or—hitting closer to present chapter—experience the fear that comes from facing someone as ruthless as Arnie. There are today no simulations of subjective consciousness (Attribute 2 in the above list of capacities constitutive of personhood). There *are* robots that encourage humans seeing them to ascribe consciousness to them. But that is quite different, as explained in (Bringsjord 2007).

Notice that while you may disagree about what is *ultimately* mechanizable, you must concede that, at least as of *now*, we have no third-person formalization of consciousness. In other words, property dualism (the view that such properties as "experiencing the deep fear that comes from facing such a one as Arnie" are incorporeal, though they may be *correlated* with certain neurophysical properties of the brain) may be false, but it currently can't be overthrown by providing the third-person description of the physical properties that are identical to the psychological (or, as they are sometimes called in philosophy, "Cartesian"") ones. Tied to the scenario of Fenwick and Arnie with which we began, the point is that while, as we've seen, we can look inside Arnie's mind and find the structures constitutive of his belief that Muhammad will fold if subjected to exceedingly harsh tactics, we can't find anything constitutive of Arnie's enjoying the pain that his interrogation inflicts. (Recall what Francis Sullivan tells Fenwick about the limits of brain imaging technology.) As we have seen, we can formalize propositional attitudes like *believing* and *knowing*; we cannot, alas, do any such thing for emotional states. Of course, though these states are not part of the technology we envision, it is far from clear that Fenwick Gibsen is thereby hampered. When he imagined his characters before him in his home, it made no difference that the apparitions failed to literally have conscious states. What was important, and what is important in the remarkable home office that he subsequently built, is that the characters were real *to him*, because that, in the end, is all it takes for Fenwick Gibsen (and indeed for any real-life creative writer using the synthetic worlds technology we envision) to be a better dramatist.

References

Anderson, J. and Lebiere, C. 2003. The Newell test for a theory of cognition. *Behavioral and Brain Sciences,* 26, 587–640.

Arkoudas, K. & Bringsjord, S. 2008. Toward formalizing common-sense psychology: an analysis of the false-belief task, in *Proceedings of the Tenth Pacific Rim International Conference on Artificial Intelligence* (PRICAI 2008) *Lecture Notes in Artificial Intelligence* (LNAI), No. 5351, edited by T.-B. Ho and Z.-H. Zhou. New York, NY: Springer-Verlag, 17–29. Preprint available at http://kryten.mm.rpi.edu/CognitiveCalculus092808.pdf [accessed 18 March 2009].

Bringsjord, S. 2008a. The logicist manifesto: at long last, let logic-based artificial intelligence become a field unto itself. *Journal of Applied Logic,* 6(4), 502–525. Preprint available at: http://kryten.mm.rpi.edu/SB_LAI_Manifesto_091808.pdf [accessed 18 March 2009].

Bringsjord, S. 2008b. Declarative/logic-based cognitive modeling, in *The Handbook of Computational Psychology,* edited by R. Sun. Cambridge, UK: Cambridge University Press, 127–169. Preprint available at http://kryten.mm.rpi.edu/sb_lccm_ab-toc_031607.pdf. [accessed 18 March 2009].

Bringsjord, S., Taylor, J., Shilliday, A, Clark, M. & Arkoudas, K. 2008. Slate: an argument-centered intelligent assistant to human reasoners, in *Proceedings of the 8th International Workshop on Computational Models of Natural Argument* (CMNA 8), edited by F. Grasso, N. Green, R. Kibble, C. Reed, 1–10, Patras, Greece, July 21, ISBN: 978-960-6843-12-9. Offprint available at http://kryten.mm.rpi.edu/Bringsjord_etal_Slate_cmna_crc_061708.pdf [accessed 18 March 2009].

Bringsjord, S., Shilliday, A.,Taylor, J., Werner, D., Clark, M., Charpentier, E. & Bringsjord, A. 2008. Toward logic-based cognitively robust synthetic characters in digital environments, in *Artificial General Intelligence 2008. Frontiers in Artificial Intelligence and Applications, edited by P.* Wang, B. Goertzel, S. Franklin, S. Amsterdam, The Netherlands: IOS Press, 87–98. Available at: http://kryten.mm.rpi.edu/sb_etal_agi2008_frames.pdf [accessed 18 March 2009].

Bringsjord, S., Arkoudas, K., Clark, M., Shilliday, A., Taylor, J., Schimanski, B. & Yang, Y. 2007. Reporting on some logic-based machine reading research, in *Proceedings of the 2007 AAAI Spring Symposium: Machine Reading* (SS–07–06. Menlo Park, CA: AAAI Press, 23–28. Available at: http://kryten.mm.rpi.edu/sb_ka_machinereading_ss07_012907.pdf [accessed 18 March 2009].

Bringsjord, S. 2007. Offer: one billion dollars for a conscious robot. If you're honest, you must decline. *Journal of Consciousness Studies*, 14(7), 28–43. Offprint available at http://kryten.mm.rpi.edu/jcsonebillion2.pdf [accessed 18 March 2009].

Bringsjord, S. & Yang, Y. 2003. Representations using formal logics, in *Encyclopedia of Cognitive Science Vol. 3*, edited by L. Nadel. London, UK: Nature Publishing Group, 940–950. Offprint available at http://kryten.mm.rpi.edu/SBYY.ECS.pdf [accessed 18 March 2009].

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

Bringsjord, S. 2000. A contrarian future for minds and machines. *Chronicle of Higher Education*, November 3, B5. (Reprinted in *The Education Digest* 66(6), 31–33.)

Bringsjord, S. 1999. The zombie attack on the computational conception of mind. *Philosophy and Phenomenological Research*, 59(1), 41–69.

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

Bringsjord, S. 1995. Pourquoi Hendrik Ibsen Est-Is Une Menace Pour La Littérature Générée Par Ordinateur?. (traduit par Michel Lenoble) in *Littérature et Informatique la Littérature Générée Ordinateur*, edited by A. Vuillemin. Arras, France: Artois Presses Universite), 135–144.

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

Charniak, E. 1993. *Statistical Language Learning*. Cambridge, MA: MIT Press.

Charniak, E. and McDermott, D. 1985. *Introduction to Artificial Intelligence*. *Reading*, MA: Addison-Wesley.

Chisholm, R. 1976. *Person and Object: A Metaphysical Study*. London, UK: George Allen and Unwin.

Dennett, D. 1978. Conditions of personhood, in *Brainstorms: Philosophical Essays on Mind and Psychology*. Montgomery, VT: Bradford Books, 267–285.

Fagin, R., Halpern, J., Moses, Y. & Vardi, M. 2004. *Reasoning About Knowledge*. Cambridge, MA: MIT Press.

Fuchs, N., Schwertel, U. & Schwitter, R. 1999. *Attempto Controlled English (ACE) Language Manual, Version 3.0*. No. 99.03. Zurich, Switzerland: Department of Computer Science, University of Zurich.

Kamp, H. & Reyle, U. 1996. A calculus for first order discourse representation structures. *Journal of Logic, Language and Information,* 5, 297–348.

Kamp, H. & Reyle, U. 1993. *From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. First edition. New York, NY: Springer.

Manzano, M. 1996. *Extensions of First Order Logic*. Cambridge, UK: Cambridge University Press.

Moravec, H. 1999. *Robot: Mere Machine to Transcendent Mind*. Oxford, UK: Oxford University Press.

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

Premack, D. & Woodruff, G. 1978. Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences,* 4, 515–226.

Russell, S. & Norvig, P. 2002. *Artificial Intelligence: A Modern Approach*. Second edition. Upper Saddle River, NJ: Prentice Hall.

Turing, A. 1950. Computing machinery and intelligence. *Mind* LIX, 59(236), 433–460.

Wiener, P. 1951. *Leibniz: Selections*. *New* York, NY: Scribner.

Wimmer, H. & Perner, J. 1983. Beliefs about beliefs: representation and constraining function of wrong beliefs in young children’s understanding of deception. *Cognition,* 13, 103–128.

Yang, Y. & Bringsjord, S. 2003. Newell's program, like Hilbert’s, is dead; let's move on. *Behavioral and Brain Sciences,* 26(5), 627.