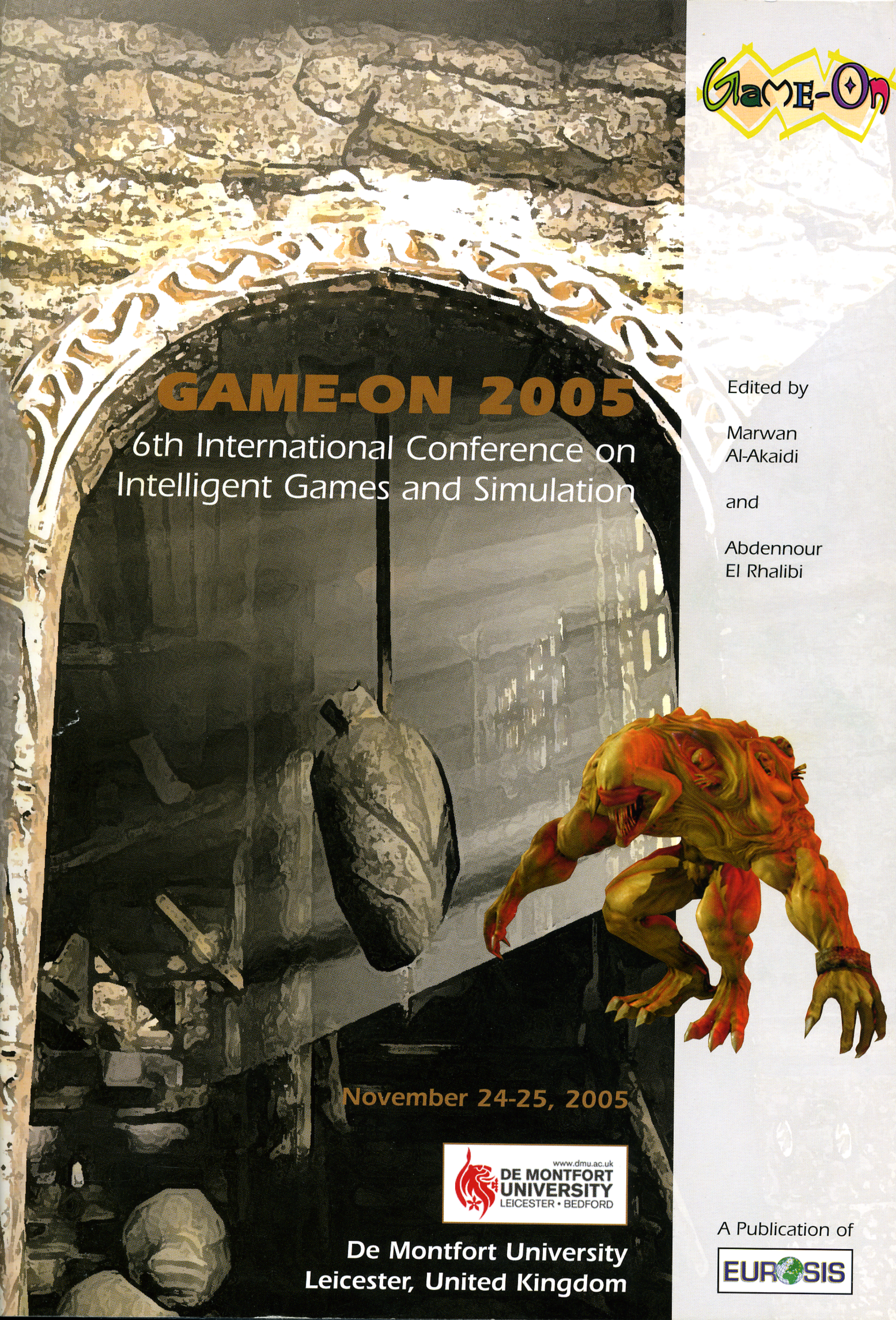

GAME-ON 2005

6th International Conference on Intelligent Games and Simulation

Edited by

Marwan Al-Akaidi

and

Abdennour El Rhalibi

November 24-25, 2005

www.dmu.ac.uk
DE MONTFORT UNIVERSITY
LEICESTER • BEDFORD

De Montfort University
Leicester, United Kingdom

A Publication of EUROSIS

Advanced Synthetic Characters, Evil, and E*

Selmer Bringsjord[1], Sangeet Khemlani[2], Konstantine Arkoudas[3], Chris McEvoy[4], Marc Destefano[5], Matthew Daigle[6]
Department of Cognitive Science[1−5]
Department of Computer Science[1,3,4]
Rensselaer AI & Reasoning Laboratory:[1−5]
http://www.cogsci.rpi.edu/research/rair/index.php
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA
{selmer,arkouk,mcevoc,khemls,destem}@rpi.edu
6: Dept. of Computer Science Vanderbilt University Nashville TN mdaigle@isis.vanderbilt.edu

Abstract

We describe our approach to building advanced synthetic characters, within the paradigm of logic-based AI. Such characters don't merely evoke *beliefs* that they have various mental properties; rather, they must actually *have* such properties. You might (e.g.) believe a standard synthetic character to be evil, but you would of course be wrong. An *advanced* synthetic character, however, can literally *be* evil, because it has the requisite desires, beliefs, and cognitive powers. Our approach is based on our RASCALS architecture, which uses simple logical systems (first-order ones) for low-level (perception & action) and mid-level cognition, and advanced logical systems (e.g., epistemic and deontic logics) for more abstract cognition. To focus our approach herein, we provide a glimpse of our attempt to bring to life one particular advanced synthetic character from the "dark side" — the evil character known simply as E. Building E entails that, among other things, we formulate an underlying logico-mathematical definition of evil, and that we manage to engineer both an appropriate presentation of E, and communication between E and humans. For presentation, which we only encapsulate here, we use several techniques, including muscle simulation in graphics hardware and approximation of subsurface scattering. For communication, we use our own new "proof-based" approach to Natural Language Generation (NLG). We provide an account of this approach.

The Dearth of AI in AI

There's an unkind joke — which made the rounds (e.g.) at the Fall 2004 AAAI Fall Symposium on Human-Level AI — about the need to create, within AI, a special interest group called 'AI'. This kind of cynicism springs from the not uncommon, and not totally inaccurate, perception that most of AI research is aimed at exceedingly narrow problems light years away from the cognitive capacities that distinguish human persons.[1]

Human-level AI is now so unusual that an entire upcoming issue of *AI Magazine* will be devoted to the subject — a bit odd, given that, at least when the field was young, AI's journal of record would have *routinely* carried papers on mechanizing aspects of human-level cognition. Seminal AI thinkers like Simon, Newell, Turing — these researchers didn't shy away from fighting to capture human-level intelligence in machine terms. But now their attitude seems moribund.

But gaming, simulation, and digital entertainment (and hereafter we refer simply to 'gaming' to cover this entire field/market), thankfully, are different: *ultimately* anyway, they call for at least the *appearance* of human-level AI (Bringsjord 2001). (On a case-by-case basis, as various games show (e.g., *The Sims* (Electronic Arts Inc. 2000)), a *non*-advanced character will of course do just fine.) Gaming doesn't strive just for a better SAT-based planner, or another tweak in a learning algorithm that doesn't relate in the least to human learning. A SAT planner doesn't constitute a virtual person. But that's precisely what we want in gaming, at least ultimately. And even in the short term we want characters that at least *seem* human. Methodologically speaking, gaming's best bet for characters that seem human is to bite the bullet and strive to engineer characters that have what it takes to *be* human. This, at least, is our strategy.

Gaming and Full-Blown Personhood

Now, there are various ways to get clearer about what gaming, at least in the long-term, needs when it comes to human-level intelligence. One way is to say simply that gaming needs artificial creatures which, behaviorally at any rate, satisfy one or more plausible proposed definitions of personhood in the literature. One such definition has been proposed by Bringsjord in (Bringsjord 1997). This definition essentially amounts to the view that x is a person if and only if x has the *capacity*

1. to "will," to make choices and decisions, set plans and projects — autonomously;
2. for consciousness, for experiencing pain and sorrow and happiness, and a thousand other emotions — love, passion, gratitude, and so on;
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, and to believe things about what others believe about one's beliefs (and so on);
6. to desire not only particular objects and events, but also changes in his or her character;

*The R&D described in this paper has been supported in part by much appreciated grants from AFRL-Rome and DARPA-IPTO.

[1] An endless source of confirming examples can be found in the pages of the *Machine Learning* journal. The dominant learning technique that you yourself employ in striving to learn is *reading*; witness what you're doing at the moment. Yet, a vanishingly small amount of R&D on learning is devoted to getting a computer program to learn by reading.

7. to reason (for example, in the fashion exhibited in the writing and reading of this very paper).

Unfortunately, this list is daunting, especially if, like us, you really and truly want to engineer a virtual person in the *short term*. A large part of the problem is consciousness, which we still don't know how to represent in third-person machine terms (Bringsjord 1998; Bringsjord 2001). But even if we leave aside consciousness, the rest of the attributes in the above list make for mighty tough challenges. In the section *"Making the Challenge of Personhood Tractable"* we shall retreat from this list to someting doable in the near term, guided by particular scenarios that make natural homes for E. But in the end, whatever appears on this list is an engineering target for us; in the long term we must confront each clause. Accordingly, in the section *"How Does E Talk?"* we explain how we are shooting for clause 4, communication. We have made progress on some of the other clauses, but there is insufficient space to present that progress herein. Clause 5 is one we believe we have pretty much satisfied, via the formalization and implementation given in (Arkoudas & Bringsjord 2005).[2]

Current State of the Art versus Computational Persons

Synthetic Characters in Gaming

What's being done now in gaming, relative to full-blown personhood, is clearly inadequate; this can be quickly seen by turning to some standard work: Figure 1 shows an array of synthetic characters from the gaming domain; these will be familiar to many readers.[3]

None of these creatures has anything close to the distinguishing features of personhood. Sustained treatments of synthetic characters and how to build them are similarly limited. For example, consider Figure 2, taken from (Champandard 2003).[4] As a mere FSA, there is no knowledge and belief, no reasoning, no declarative memories, and no linguistic capacity. In short, and this is perhaps a better way of putting the overall problem infecting todays's virtual characters, all of the cognitive capacities that distinguish human persons, according to the science of cognition (e.g., (Goldstein 2005)), are missing. Even the state of the art using cognitive architectures (e.g., SOAR) is primitive when stacked against full-blown personhood (Ritter *et al.* June 2002).

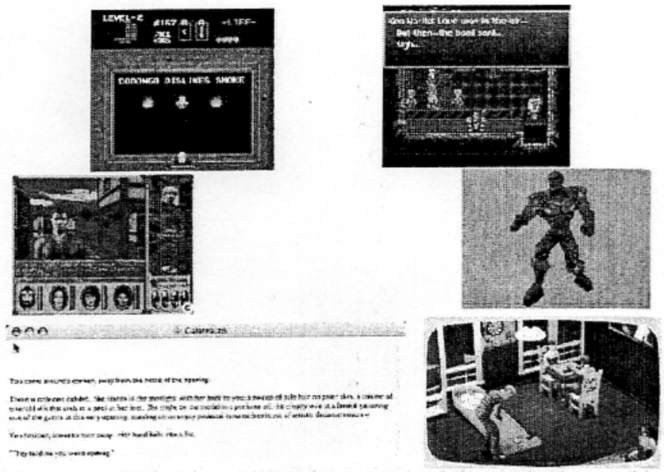

Figure 1: Sample Synthetic Characters

What About Synthetic Characters in Cutting Edge Research?

What about research-grade work on synthetic characters? Many researchers are working on synthetic characters, and have produced some truly impressive systems. However, all such systems, however much they *appear* to be human persons, aren't. We now consider three examples of such work, and show in each that the character architectures don't have the underlying cognitive content that is necessary for personhood.

REA

An agent developed by (Cassell *et al.* 1999) known as REA is an example of a successful, robust agent whose developers focused primarily on embodied conversation and the conversational interface. She is described as being an expert in the domain of real estate, and interactions with REA are both believable and informative.

REA, however, is representative of many of the industry's most successful agents in that she excels at content management, but fails to deliver rich emotive and cognitive functionality. REA, after all, cannot generate English from *arbitrary* underlying knowledge. Like many of her peers, REA's underlying cognitive capabilities are modeled in an ad-hoc fashion. Her personality is in no way defined; her interactions within a particular situation lack subtlety and depth. While she excels as a simulated character and a conversational agent, she is bereft of the rich cognitive content with which advanced synthetic characters must behave.

The BEAT Architecture

In an engaging paper (Gratch *et al.* 2002), Gratch and colleagues present an architecture for developing rich synthetic characters. This architecture is known as the Behavior Expression Animation Toolkit Text-to-Nonverbal Behavior Module (BEAT). Under this architecture, emotion and cognitive content are produced systematically in a *simulation-based* approach.

Their simulation-based approach is built on top of *ap-*

[2] A preprint is available online at http://kryten.mm.rpi.edu/arkoudas.bringsjord.clima.crc.pdf.

[3] Worst to best, in our eyes: Top-left, The Legend of Zelda; SC spits text upon entering room. Top-right, Chrono Trigger; tree-branching conversations. Middle-left, Might & Magic VI (Shopkeepers). Middle-right, Superfly Johnson from Daikatana; behavior scripting, attempts to follow player and act as a sidekick (fails!). Bottom-left, Galatea – Interactive Fiction award winner for Best NPC of 2000 (text-based). Bottom-right, Sims 2. But even here, nothing like what our RASCALS architecture has is present.

[4] This is an excellent book, and it's used in our lab for building synthetic characters. But relative to the loftier goals of reaching *bona fide* personhood in artificial characters, there's clearly a lot of work to be done.

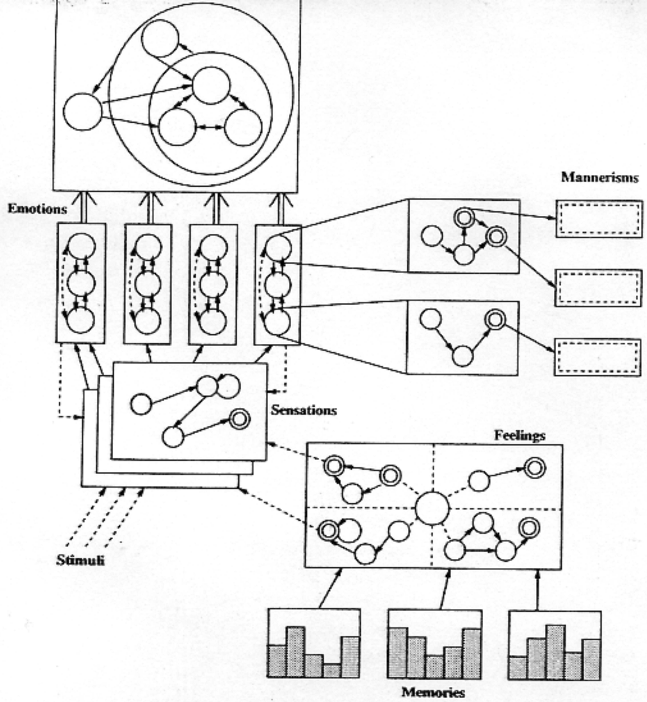

Figure 2: Impoverished Formalism for Synthetic Characters

praisal theories of emotion, where emotions emerge from analysis of events and objects in a particular domain with respect to the agent's goals, standards, and attitudes. But as Gratch et al themselves point out, appraisal theories "are rather vague about the assessment process...A promising line of research is integrating AI-based planning approaches, which might lead to a concretization of such theories." We will present the RASCALS paradigm as one that utilizes precisely the AI-based planning techniques Gratch et al. regard as promising.

Unfortunately, while Gratch and colleagues make wonderful advancements in the logistics of realizing agents, the issue of developing rich underlying cognitive content is eschewed. Even assuming that their simulation-based approach utilizes robust AI-based planning, the focus is not on developing true cognitive content but rather on its simulation and modeling.

Believable Interactive Embodied Agents

An approach more focused on building believable characters was proposed by (Pelachaud & Poggi 2002). They argue that research should include three distinct phases:

- *Phase 1: Empirical Research.* This phase involves research "aimed at finding out the regularities in the mind and behavior of Human Agents, and at constructing models of them."
- *Phase 2: Modeling Believable Interactive Embodied Agents.* Here, "rules are formalized, represented, and implemented in the construction of Agents."
- *Phase 3: Evaluation.* Finally, agents are tested on several levels, including "how well they fit the User's needs and how similar they look to a real Human Agent."

The "rule formalization" characterized in Phase 2 is, as Pelachaud and Poggi point out, indispensable when building believable characters. Since such rule formalizations are all expressible in first-order logic, their approach is actually a proper subset of the RASCALS approach. But formalizing and implementing rules is not enough to achieve true cognition; after all, cognition involves much more than simple rules/first-order logic. Iterated beliefs are beyond the reach of first-order logic. Finally, while Pelachaud and Poggi elaborate on linguistic rules and formalizations, they fail to mention anything about modeling cognition or interacting with a given knowledge base, and they make no remarks concerning the logistics behind rule formalization and implementation. The agents described therein all possess rudimentary cognitive content but come nowhere close to true cognitive or emotive capacity.

Making the Challenge of Personhood Tractable

How can we make the challenge of engineering a virtual person tractable in the very short term? Our lab has a two-part answer. First, assimilate everything out there regarding the *craft* of making viewers and users *believe* that the synthetic character they interact with is a genuine person. This is the same route that was followed by Bringsjord and Ferrucci in the design of the BRUTUS story generation system (Bringsjord & Ferrucci 2000). In a nutshell, B&F studied the literature on what responses are desired in readers by clever authors, and then reverse engineered back from these responses to a story generation system that triggers some of them. In connection with synthetic characters, this general strategy has impelled us to build up a large library on the design of synthetic charaters in stories and movies. In addition, we have built up a library of characters in film — specifically one that specializes in candidates for true evil. Within the space we have herein, however, this general strategy, and the results so far obtained, can't be presented. So we will settle here for a shortcut; it's the second part of our two-part answer. The shortcut is to work from concrete scenarios backwards by reverse engineering. We currently have two detailed scenarios under development. One is based on the evil people whose personalities are revealed in conversations in (Peck 1983); we leave this one aside for now. The second scenario, which is part of R&D undertaken in the area of wargaming, can be summarized as follows. (At the conference, we would provide a demo of conversation with E regarding the first of these scenarios, where that conversation conforms to our account of evil; see *On our Formal Account of Evil*.)

E in Scenario 2, and Inference Therefrom

Let us imagine a man named simply E, a brutal warlord in a war-torn country. E is someone you're going to have to vanquish. He has moved up the ranks of the underworld in post-apocalyptic America after "success" in many, many murderous missions. E has taken a number of prisoners from an organization (let's call it simply O) he seeks to intimidate. O is chosen specifically because it is trying to rebuild the fractured US in the direction of a new federal governing[5]. Conforming to what has unfortunately become a gruesome pattern, E decides to film the beheading of one of these poor prisoners, and to release the video to O.

[5]Coincidentally, we have recently learned that the game *Shattered World* for the XBox is related to our scenario.

Given just this small amount of information, what can we infer about E's knowledge and reasoning? That it has at least the following six attributes:

1. *Mixed Representation.* E's knowledge is not simply linguistic or symbolic in nature. It includes visual or pictorial knowledge as well. For example, E clearly is thinking in terms of mental images, because he plans to gain leverage from the release of images and video. In addition, though it isn't pleasant to contemplate, E certainly has a "mental movie" that he knows he can turn into real life: he envisions how such executions work before performing them.
2. *Tapestried.* Presumably E's knowledge of his prisoners is relatively new. But this new knowledge is woven together with extensive prior knowledge and belief. For example, in E's case, he has extensive knowledge of O, and its principles regarding treatment of prisoners.
3. *Extreme Expressivity.* E's knowledge and reasoning requires highly expressive propositions. For example, he believes that O believes that it is universally forbidden to execute prisoners, and he believes that some of those aiding the United States' rebuilding effort will be struck with fear once the execution is complete and suitably publicized, and that that fear will affect their beliefs about what they should and shouldn't do.
4. *Mixed Inference Types.* E's reasoning is based not only on deductive inference, but also on educated guesses (abduction), and probabilistic inference (induction).
5. *Uses Natural Language.* E communicates in natural language, with his comrades, and with others as well.
6. *Multi-Agent Reasoning.* E is of course working in coordinated fashion with a number of accomplices, and to be effective, they must reason well as a group.

Working within the paradigm of logic-based AI (Bringsjord & Ferrucci 1998a; Bringsjord & Ferrucci 1998b; Nilsson 1991; Genesereth & Nilsson 1987), and using the MARMML knowledge representation and reasoning system, which is based on: the theory known as mental metalogic (Yang & Johnson-Laird 2000a; Yang & Johnson-Laird 2000b; Yang & Bringsjord 2005; Rinella, Bringsjord, & Yang 2001; Yang & Bringsjord 2001a; Yang & Bringsjord 2001b; Yang, Braine, & O'Brien 1998), the Denotational Proof Language known as Athena (Arkoudas 2000), Barwisean grids for diagrammatic knowledge and reasoning (see the mathematical section of (Barwise & Etchemendy 1995)), and RASCALS[6](see Figure 3), a revolutionary architecture for synthetic characters, we are building a virtual version of E that has the six attributes above.

Brief Remarks on the RASCALS Architecture

Let us say a few words about RASCALS, a brand new entry in the field of compuational cognitive modeling, which revolves around what are called *cognitive architectures* (e.g., SOAR (Rosenbloom, Laird, & Newell 1993); ACT-R (Anderson 1993; Anderson & Lebiere 1998; Anderson & Lebiere 2003); CLARION (Sun 2001); Polyscheme (Cassimatis 2002; Cassimatis *et al.* 2004)). What makes the RASCALS cognitive architecture distinctive? There is insufficient space here to convey any technical detail (for more details, see (Bringsjord forthcoming)); we make just three quick points about RASCALS, to wit:

[6]**R**ensselaer **A**dvanced **S**ynthetic **C**haracter **A**rchitecture for **L**ogical **S**ystems

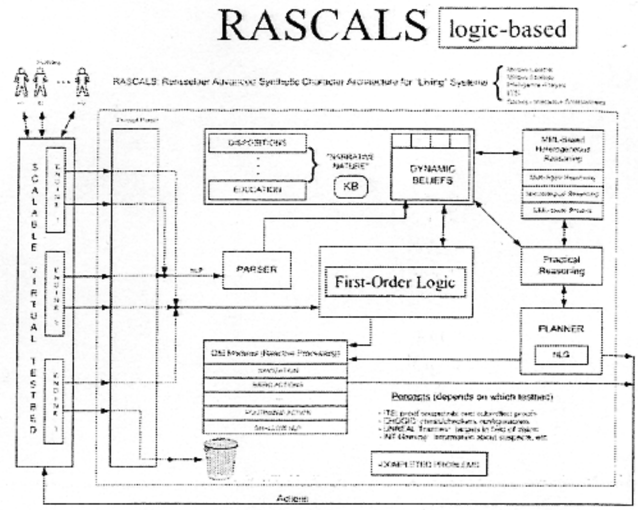

Figure 3: RASCALS: **R**ensselaer **A**dvanced **S**ynthetic **C**haracter **A**rchitecture for **L**ogical **S**ystems

- All other cognitive architectures we know of fall far short of the expressive power of RASCALS. For example, SOAR and ACT-R struggle to represent (let alone reason quickly over) textbook problems in logic (e.g., the Wise Man Problem = WMP) but in RASCALS such representations are effortless (see (Arkoudas & Bringsjord 2005) for the solution to WMP in Athena, included in RASCALS).
- The great challenge driving the field of computational cognitive modeling (CCM) is to unify *all* of human cognition; this challenge can be traced back to the birth of CCM in the work of Newell 1973. Such unification is achieved in one fell swoop by RASCALS, because *all* of cognition can be formalized and mechanized in logic (though doing so requires some very complicated logics well beyond first-order logic, as in (Arkoudas & Bringsjord 2005)).[7]
- While logic has been criticized as too slow for real-time perception-and-action-heavy computation, as you might see in first-person shooter (as opposed to a strategy game, which for obvious reasons fits nicely with the paradigm of logic-based AI), it has been shown that RASCALS is so fast that it can enable the real-time behavior of a mobile robot. We have shown this by having a logic-based mobile robot successfully navigate the wumpus world game, a

[7]It will naturally occur to some skeptics to enquire about traditional-style learning, and speech recognition. As to the former, it's well-known that there are logic-based approaches to divining a function f by repeated trial; see, e.g., (Russell & Norvig 2002). There are also well-known knowledge-based (which become, in RASCALS, more formal, logic-based) techniques for learning: EBL, RBL, etc.; again, see (Russell & Norvig 2002) for a survey. Of course, RASCALS does reject purely statistical and probabilistic approaches to learning (and other cognitive phenomena). That seems quite unsurprising, since statistical approaches in AI routinely reject, to their peril, declarative/logic-based techniques. As to the latter problem, RASCALS insists that all language be represented in logical form, and Bringsjord concedes that this is currently not achieved, nor even on the horizen. However, with respect to natural language understanding, all researchers, whatever their approach, are currently in the same dismal boat.

staple in AI. (See Figures 4 and 5.)

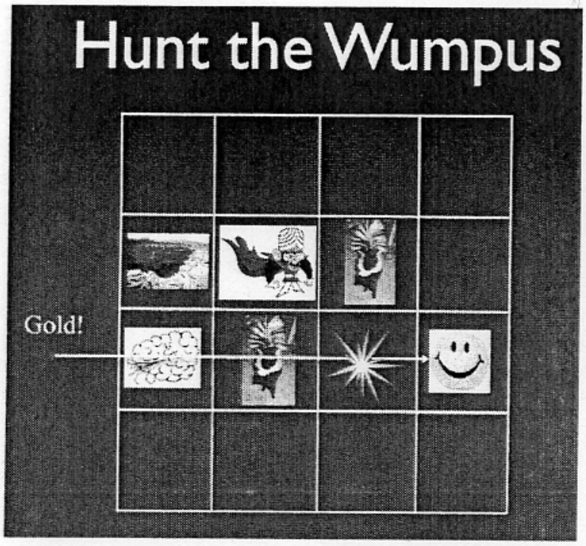

Figure 4: The Wumpus World Game

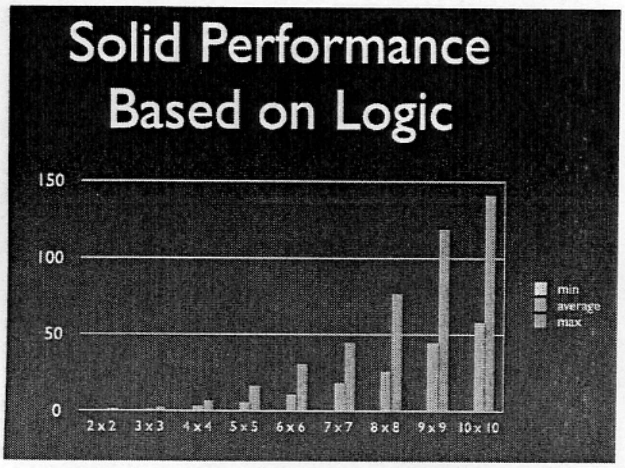

Figure 5: Performance of a RASCALS-Powered Robot in the Wumpus World

To show part of the underlying structure of E in connection with the attribute *Extreme Expressivity*, we now present an informal version of the formal account of evil that is implemented in our RASCALS architecture. This account specifically requires logics expressive enough to handle knowledge, belief, and ethical concepts. These logics go well beyond first-order logic; details and an implementation can be found in (Arkoudas & Bringsjord 2005). In the section *"E: The Presentation Level"* we explain the technology that allows E to speak naturally in English; that is, we show there part of the underlying structure of E associated with *Uses Natural Language*.

On our Formal Account of Evil

If we charitably push things in the direction of formally representing a definition of evil,[8] then we can understand Feinberg 2003 as advancing pretty much this definition:

Def 1 Person s is evil iff there exists some action a[9] such that

1. performing a is morally wrong;
2. s is morally blameworthy for performing a;
3. s's performing a causes considerable harm to others; and
4. the reasons or motives for s's performing a, along with "the elements that ground her moral blameworthiness," are unintelligible.

This is a decent starting place, but for us there are problems. For example, imagine that E invariably fails to cause *actual* harm. Surely he would still qualify as evil even if he were a bumbling villain. (If the knife slipped when he attempted decapitation, he would still be just as black-hearted.) This means that clause 3 should at least be replaced by

$3'$. s performs a in the hopes of causing considerable harm to others

But even this new definition, for reasons we don't have space to explain, is wholly inadequate. To give just a flavor for what E is currently based upon, we present simply our current best replacement for clause 4:

$4''$ were s a willing and open participant in the analysis of reasons and motives for s's seeking to perform a, it would be revealed that either

(i) these reasons and motives are unintelligible, or

(ii) s seeks to perform a in the service of goal g, and

(a) the anticipatable side-effects e of performing a are bad, but s cannot grasp this, or

(b) g itself is appraised as good by s when it is in fact bad.

Just this clause alone required much sustained analysis. (For a full chronicle of the evolution of a formally refined definition of betrayal from a rough starting one, see the chapter "Betrayal" in (Bringsjord & Ferrucci 2000).)

Keep in mind that this is still informal, kept that way in the interests of easing exposition. In the RASCALS-based implementation of E, evil must be expressed in purely formal form, which requires, again, that we use advanced logics of belief, knowledge, and obligation.[10]

Keep in mind as well that we're not claiming that we have the perfect definition of evil. Some may object to our definition, and some of their objections may be trenchant. But the important point is to see how rich evil is — to see that it involves all kinds of highly cognitive powers and concepts that simply aren't found in today's synthetic characters. To be evil, one has to have beliefs, desires, and one has to have a lot of knowledge. The detailed configuration of these elements may not be exactly as we claim they ought to be, but no one can deny that the elements are needed. Without them, a synthetic character who is supposed to be evil is only a fake shell. And in the end, the shell will be revealed to be a shell: the illusion, at some point, will break down.

[8]Feinberg's work is informal, and not suitable for direct use in AI and computer science.

[9]Or omission.

[10]For a look at the deontic logic (i.e., the logic of ethical concepts) we are relying upon, see (Horty 2001). Our mechanization of this this logic will be presented at the AAAI November 2005 Fall Symposium on Machine Ethics. The paper is available online at http://kryten.mm.rpi.edu/FS605ArkoudasAndBringsjord.pdf.

How Does E Talk?

As everyone knows, once the daunting challenge of rendering consciousness in computational terms is put aside, the greatest remaining challenge is that of giving an advanced synthetic character the power to communicate in a natural language (English, French, etc.) at the level of a human person. As you'll recall, communicative capacity is one of the clauses in the definition of personhood presented above. A plausible synthetic character must necessarily communicate in a fluid, robust manner. How, then, is such a rich form of communication implemented in E?

Reconciling Knowledge Representation and NLG

E speaks by parsing and processing formal knowledge; he develops an ontology based on internal and external queries, and then reasons over his knowledge to produce meaningful content. This content is then sent to his NLG module, translated into English, and finally presented to the user. Before we examine what goes on inside E's NLG module, let's take a moment to examine how E produces "meaningful content."

When we ask E a question, we are clearly interested in an answer that is both relevant and meaningful, an answer indistinguishable from those given by a real person. Assuming we have incomplete knowledge, suppose we ask of E, "Is John dangerous?" E approaches this question through formal logical analysis. The idea is to have E determine incontrovertibly whether John is dangerous or not. So, for instance, suppose E's knowledge base includes the following three facts:

1. DANGEROUS PEOPLE HAVE AUTOMATIC WEAPONS.
2. JOHN HAS A BERETTA AR-70 ASSAULT RIFLE.
3. THE BERETTA AR-70 ASSAULT RIFLE IS AN AUTOMATIC WEAPON.

None of the information above explicitly tells E whether John is dangerous or not, but clearly, when presented the above query, we want E to answer with an emphatic "Yes." Still, the answer itself is not enough. To ensure that E understands the nature of the question as well as the information he is dealing with, he must, upon request, provide *a justification for every answer.* The justification presented to the user is a formal proof, translated into English. Thus, E could answer as follows:

JOHN IS IN FACT DANGEROUS BECAUSE HE HAS A BERETTA AR-70 ASSAULT RIFLE. SINCE A BERETTA AR-70 ASSAULT RIFLE IS AN AUTOMATIC WEAPON, AND SINCE DANGEROUS PEOPLE HAVE AUTOMATIC WEAPONS, IT FOLLOWS THAT JOHN IS DANGEROUS.

Content is thus generated in the form of a formal proof. In general, the proofs generated will be more complex (they will use larger knowledge bases) and more sophisticated (they will use deontic and epistemic logic).

While the example is simple and rudimentary (that is, it makes use of only first-order logic and a small knowledge base), it demonstrates that E is taking heed of his knowledge to generate a meaningful reply. In the RASCALS architecture, answering "Yes" to the query above implies that E must in fact have the corresponding knowledge, an implication that does not hold for other architectures.

For a more formal method of analysis, we introduce the "Knowledge Code Test": If synthetic character C says something X or does something X designed to evoke in the mind of the human gamer/user the belief that C knows $P_1, P_2, \ldots$, then we should find a list of formulas, or the equivalent, corresponding to $P_1, P_2, \ldots$ in the code itself. The characters in Figure 1 would fail such a test, as would characters built on the basis of Champandard's specifications. An FSA, as a matter of mathematical fact, has no storage capability. A system with power that matches that of a full Turing machine is needed to pass the Knowledge Code Test (Lewis & Papadimitriou 1981).

But formal proofs are oftentimes too detailed to be of interest. Before we can even begin translating a proof into an English justification, we need verify that its level of abstraction is high enough that it is easy to read and understand. After all, formal natural deduction proofs are difficult and tedious to read. To represent proofs at a more wholistic, abstract level, we utilize the denotational proof language known as Athena (Arkoudas 2000). Athena is a programming language, development environment, and interactive proof system that evaluates and processes proofs as input. Its most prominent feature is its ability to present proofs in an abstract, top-level manner, isomorphic to that of a natural argument a human might use. By developing proofs in Athena at this level, a level high enough to be of interest to a human reader, we can be sure that the language generated from our NLG module is at precisely the level of abstraction we desire — neither too detailed nor too amorphous.

It's now time to look at precisely how English is generated from a formal proof.

Proof-based Natural Language Generation

Very few researchers are experimenting with the rigorous translation of formal proofs into natural language[11]. This is particularly odd when one considers the benefits of such a program. Natural deduction proofs, provided that they are developed in a sensible manner, are already poised for efficient translation. They require absolutely *no* further document structuring or content determination. That is, document planning, as defined by (Reiter & Dale 2000), is completely taken care of by using formal proofs in the first place.

Our NLG module receives as input a formal proof and returns as output English text. The English generated is an isomorph of the proof received. The structure of the justification, then, is precisely the same as the structure of the proof. If the justification uses *reductio ad absurdum* in the middle of the exposition, then you can be sure that there's a proof by contradiction in the middle of the formal proof.

Formal proofs are constructed from various different subproofs. A proof by contradiction is one such example of a type of subproof, but there are of course many others. Our system breaks a proof down to its constituent subproofs, translating each subproof from the top down. For example, assume the following:

1. CHICAGO IS A TARGET OR NEW YORK IS A TARGET

[11]An example of one such team is a research group at the University of Saarlande. The group had, at least until 1997, been developing a system called PROVERB (Huang & Fiedler 1997). Their approach to proof-based translation was unique and extremely influential, though their project was largely unsuccessful.

2. IF CHICAGO IS A TARGET, MILLIONS WILL DIE.
3. IF NEW YORK IS A TARGET, MILLIONS WILL DIE.

To deduce something meaningful from this information, we'll use a proof by cases. Our system translates this proof form as follows:

RECALL THAT CHICAGO OR NEW YORK IS A TARGET. EACH CASE PRODUCES THE SAME CONCLUSION; THAT IS, IF CHICAGO IS A TARGET THEN MILLIONS WILL DIE, AND IF NEW YORK IS A TARGET THEN MILLIONS WILL DIE. IT FOLLOWS THAT MILLIONS WILL DIE.

Predictably, documents produced in this manner, even when presented at a level abstract enough to make sense to a layperson, are rigid and, well, inhuman. They use the same phrases over and over again, they lack fluidity, and they are completely divorced of grace and wit. To boot, they disregard contextual information. Merely translating constituent subproofs to English will not produce natural English.

Nevertheless, this methodology provides a foundation for more sophisticated development. Once constituent subproofs are translated properly, they are sent to a microplanning system that maps particular subproofs to discourse relations (Hovy 1993). This mapping is known as a *message* and is not isomorphic. While the structure of the overall proof is preserved in the final document, individual subproofs are not treated with the same stringency. They can be molded and fitted to a number of different discourse relations for the sake of fluidity. Two more steps remain before natural language can be produced.

Lexicalization is the process by which a lexicon of words is selected and mapped onto its symbolic counterparts. The content implicit in the proof, structured through subproof analysis and discourse relations, needs to be lexicalized before it can be presented as English text. That is, exact words and phrases must be chosen to represent relationships and predicates. For instance, TARGET(CHICAGO) must be translated to CHICAGO IS A TARGET and BERETTA(JOHN) must be translated to JOHN HAS A BERETTA before we can move on to gluing everything together. The only way this can happen is if a lexical database such as WordNet (Miller 1995) is augmented with domain-specific lexicalizations such as those specifying how to lexicalize "Beretta AR-70."

For even more fluidity, it's necessary to avoid referring to the same entities with the same phraseology. At the very least, pronouns should be substituted when referring to repeated concepts, persons, places, and objects. These substitutions are known as *referring expressions*, and need to be generated to truly produce fluid, humanlike English.

Fortunately, once the above issues are resolved, the information gathered therein can be plugged easily into a surface realizer such as KPML (Bateman 1997). In this fashion, proof-based NLG allows for the generation of both structured and expressive expositions.

How far can an approach to NLG based on logic go? What about rhetorical structure, for example? The engineering of E reflects a belief that all of NLG, in the context of an advanced synthetic characters, can indeed be achieved through the mechanization of sufficiently complex logical systems. Only time will tell if this approach has the necessary breadth, but rhetorical structure seems particularly well-suited to capture in logic. Of note here is the fact that it was logic that dictated including the present paragraph.

E: The Presentation Level

To concretize our thoughts on evil, we show E; a realistic real-time presentation of an evil talking head in the formal sense. In order to give E a realistic look and a range facial expressions, we have created a muscle model of the face. Each simulated muscle in our model can contract and this contraction perturbs the vertices of the skin (E is rendered as a triangle mesh.) The effect of muscle simulation is supplemented by limited use of morph target based animation for some fine details. In addition, specialized actions are used to animate the eyes, jaw, and neck.

Prior Work

There have been several recent efforts in the presentation of talking heads. A VRML based approach shown by (Breton, Bouville, & Pel 2001) addresses all aspects of facial animation working in real-time, but for very low polygon models. The face is parameterized in a simplified manner similar to E.

A more complex, physics-based system is described in (Albrecht, Haber, & Seidel 2002). Here the focus is on lip synchronization and simulation of the mouth and lips. Other aspects of the face are not specifically addressed.

Our muscle simulation is based largely on that presented in (Waters 1987), (Parke & Waters 1996), and expanded upon in (Bui, Heylen, & Nijholt 2003). We chose to work from the Waters model because we feel it is most practical for real-time applications and implementation and programmable graphics hardware. We simulate two types of muscles.

Linear Muscles

Linear muscles contract along a single axis, and are parameterized by five values. The points O and T define the origin and terminus of the muscle respectively. The scalar F defines the radius from O where the effect of muscle contraction begins to decline. The angle Z defines the angle about O where the muscle affects the skin. Finally, the scalar W gives the distance between wrinkles that form as the muscle contracts.

We define a vertex shader (implemented in HLSL) that modifies the position of skin vertices based on muscle contractions. For a linear muscle this shader computes three values. Given a vertex at position P with normal N, the angular displacement, D_A, is:

$$\min(1 - (norm(P - O), norm(T - O) * C, 0)$$

C is the muscle contraction in the above equation. The radial displacement, D_R, is:

$$1 - \left[\frac{clamp(len(P-O) - F, 0, len(T-O) - F)}{(len(T-O) - F)}\right]^3$$

Finally, the wrinkle offset, D_W, is:

$$\frac{1 - fmod(len(P-O), W) - (W * 0.5))^2}{(W * 0.5)^2}$$

We combine all these values to determine the final vertex displacement:

$$(C * D_A * D_R) + (D_W * N)$$

Many of the values in the above equations can be precomputed before the vertex shader executes. The resulting implementation uses about 30 shader instructions.

Sphincter Muscles

Sphincter muscles draw together in circular shape and we use them to model the puckering lips and squinting eyes. A sphincter muscle is parameterized by an origin O, and a horizontal and vertical extent, H and V respectively. In the vertex shader the contraction of sphincter muscles displace vertices in the following manner:

$$\max\left[1 - \frac{\sqrt{((D_x^2 * V^2) + (D_y^2 * H^2))}}{H * V}, 0\right] * C$$

In the above equation, D is the vector from a skin vertex to the origin of the muscle.

Putting It All Together

Each facial expression, be it a viseme used in speech or an emotional state, is described in terms of muscle contractions. To specify those contractions and drive E's facial animation systems, we use a very simplified scripting system for triggering named expressions at specified times with specified intensity values and blending parameters. We use the data from (Bui, Heylen, & Nijholt 2004) to prevent physically impossible muscle contractions. A parameterization for the tongue similiar to (King 2001) is used. A module for eye movements implements many of the ideas presented in (Lee, Badler, & Badler 2002). Finally, we simulate subsurface scattering on the skin using the algorithm of (Sander, Gosselin, & Mitchell 2004).

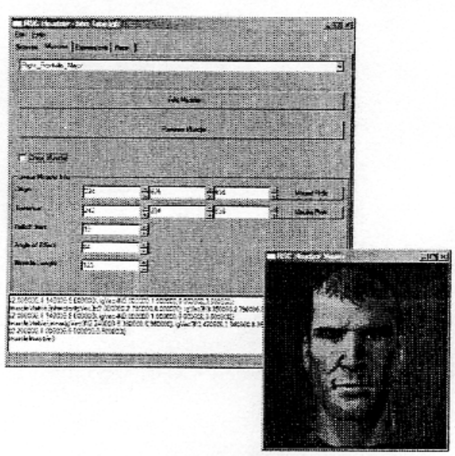

Figure 6: Tool for Manipulating Facial Muscles on E. (Note: Face shown *resembles* E's, but isn't his. E himself will be unveiled at the conference.)

Our Demos @ GameOn!

As mentioned above, at the conference we will show a conversation with E based on the first of the two aforementioned scenarios. This interaction will show our approach to the presentation level in action, and will manifest our formal account of evil in ordinary conversation that is based on our NLG technology.

References

[Albrecht, Haber, & Seidel 2002] Albrecht, I.; Haber, J.; and Seidel, H. P. 2002. Speech synchronization for physics-based facial animation. In *Proceedings WSCG*, 9–16.

[Anderson & Lebiere 1998] Anderson, J. R., and Lebiere, C. 1998. *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum.

[Anderson & Lebiere 2003] Anderson, J., and Lebiere, C. 2003. The newell test for a theory of cognition. *Behavioral and Brain Sciences* 26:587–640.

[Anderson 1993] Anderson, J. R. 1993. *Rules of Mind*. Hillsdale, NJ: Lawrence Erlbaum.

[Arkoudas & Bringsjord 2005] Arkoudas, K., and Bringsjord, S. 2005. Metareasoning for multi-agent epistemic logics. In *Fifth International Conference on Computational Logic In Multi-Agent Systems (CLIMA 2004)*, volume 3487 of *Lecture Notes in Artificial Intelligence (LNAI)*. New York: Springer-Verlag. 111–125.

[Arkoudas 2000] Arkoudas, K. 2000. *Denotational Proof Languages*. Ph.D. Dissertation, MIT.

[Barwise & Etchemendy 1995] Barwise, J., and Etchemendy, J. 1995. Heterogeneous logic. In Glasgow, J.; Narayanan, N.; and Chandrasekaran, B., eds., *Diagrammatic Reasoning: Cognitive and Computational Perspectives*. Cambridge, MA: MIT Press. 211–234.

[Bateman 1997] Bateman, J. A. 1997. Enabling technology for multilingual natural language generation: the kpml development environment. *Nat. Lang. Eng.* 3(1):15–55.

[Breton, Bouville, & Pel 2001] Breton, G.; Bouville, C.; and Pel, D. 2001. Faceengine: a 3d facial animation engine for real time. In *Proceedings of 6th International Conference on 3D Web Technology*, 15–22.

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

[Bringsjord & Ferrucci 1998b] Bringsjord, S., and Ferrucci, D. 1998b. Reply to Thayse and Glymour on logic and artificial intelligence. *Minds and Machines* 8:313–315.

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

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

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

[Bringsjord 2001] Bringsjord, S. 2001. Is it possible to build dramatically compelling interactive digital entertainment (in the form, e.g., of computer games)? *Game Studies* 1(1). This is the inaugural issue. Url: `http://www.gamestudies.org`.

[Bringsjord forthcoming] Bringsjord, S. forthcoming. The RASCALS cognitive architecture: Logic top to bottom. In Sun, R., ed., *The Handbook of Computational Cognitive Modeling*. Cambridge University Press.

[Bui, Heylen, & Nijholt 2003] Bui, T. D.; Heylen, D.; and Nijholt, A. 2003. Improvements on a simple muscle-based 3d face for realistic facial expressions. In *Proceedings of 16th International Conference on Computer Animation and Social Agents*, 33–40.

[Bui, Heylen, & Nijholt 2004] Bui, T. D.; Heylen, D.; and Nijholt, A. 2004. Combination of facial movements on a 3d talking head. In *Proceedings of Computer Graphics International*.

[Cassell *et al.* 1999] Cassell, J.; Bickmore, T.; Billinghurst, M.; Campbell, L.; Chang, K.; Vilhjalmsson, H.; and Yan, H. 1999. Embodiment in conversational interfaces: Rea. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, 520–527. New York, NY, USA: ACM Press.

[Cassimatis *et al.* 2004] Cassimatis, N.; Trafton, J.; Schultz, A.; and Bugajska, M. 2004. Integrating cognition, perception and action through mental simulation in robots. In *Proceedings of the 2004 AAAI Spring Symposium on Knowledge Representation and Ontology for Autonomous Systems*.

[Cassimatis 2002] Cassimatis, N. 2002. *Polyscheme: A Cognitive Architecture for Integrating Multiple Representation and Inference Schemes*. Ph.D. Dissertation, Massachusetts Institute of Technology (MIT).

[Champandard 2003] Champandard, A. 2003. *AI Game Development*. Berkeley, CA: New Riders.

[Electronic Arts Inc. 2000] Electronic Arts Inc. 2000. *The Sims^{TM}: The People Simulator from the Creator of SimCity^{TM}*. Austin, TX: Aspyr Media.

[Feinberg 2003] Feinberg, J. 2003. *Problems at the Roots of Law*. New York, NY: Oxford University Press.

[Genesereth & Nilsson 1987] Genesereth, M., and Nilsson, N. 1987. *Logical Foundations of Artificial Intelligence*. Los Altos, CA: Morgan Kaufmann.

[Goldstein 2005] Goldstein, E. B. 2005. *Cognitive Psychology: Connecting Mind, Research, and Everyday Experience*. Belmont, CA: Wadsworth.

[Gratch *et al.* 2002] Gratch, J.; Rickel, J.; Andre, E.; Cassell, J.; Petajan, E.; and Badler, N. 2002. Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems* 17(4):54–63.

[Horty 2001] Horty, J. 2001. *Agency and Deontic Logic*. New York, NY: Oxford University Press.

[Hovy 1993] Hovy, E. H. 1993. Automated discourse generation using discourse structure relations. *Artif. Intell.* 63(1-2):341–385.

[Huang & Fiedler 1997] Huang, X., and Fiedler, A. 1997. Proof verbalization as an application of NLG. In Pollack, M. E., ed., *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI)*, 965–970. Nagoya, Japan: Morgan Kaufmann.

[King 2001] King, S. A. 2001. *A Facial Model and Animation Techniques for Animated Speech*. Ph.D. Dissertation, Ohio State University.

[Lee, Badler, & Badler 2002] Lee, S. P.; Badler, J. B.; and Badler, N. I. 2002. Eyes alive. *ACM Transactions on Graphics* 21(3):637–644.

[Lewis & Papadimitriou 1981] Lewis, H., and Papadimitriou, C. 1981. *Elements of the Theory of Computation*. Englewood Cliffs, NJ: Prentice Hall.

[Miller 1995] Miller, G. A. 1995. Wordnet: a lexical database for english. *Commun. ACM* 38(11):39–41.

[Newell 1973] Newell, A. 1973. You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In Chase, W., ed., *Visual Information Processing*. New York: Academic Press. 283–308.

[Nilsson 1991] Nilsson, N. 1991. Logic and Artificial Intelligence. *Artificial Intelligence* 47:31–56.

[Parke & Waters 1996] Parke, F. I., and Waters, K. 1996. *Computer Facial Animation*. New York, NY: AK Peters.

[Peck 1983] Peck, M. S. 1983. *People of the Lie*. New York, NY: Simon and Shuster.

[Pelachaud & Poggi 2002] Pelachaud, C., and Poggi, I. 2002. Multimodal embodied agents. *Knowl. Eng. Rev.* 17(2):181–196.

[Reiter & Dale 2000] Reiter, E., and Dale, R. 2000. *Building natural language generation systems*. New York, NY, USA: Cambridge University Press.

[Rinella, Bringsjord, & Yang 2001] Rinella, K.; Bringsjord, S.; and Yang, Y. 2001. Efficacious logic instruction: People are not irremediably poor deductive reasoners. In Moore, J. D., and Stenning, K., eds., *Proceedings of the Twenty-Third Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates. 851–856.

[Ritter *et al.* June 2002] Ritter, F.; Shadbolt, N.; Elliman, D.; Young, R.; Gobet, F.; and Baxter, G. June 2002. Techniques for modeling human performance in synthetic environments: A supplementary review. Technical report, Human Systems Information Analysis Center, Wright-Patterson Air Force Base, OH.

[Rosenbloom, Laird, & Newell 1993] Rosenbloom, P.; Laird, J.; and Newell, A., eds. 1993. *The Soar Papers: Research on Integrated Intelligence*. Cambridge, MA: MIT Press.

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

[Sander, Gosselin, & Mitchell 2004] Sander, P. V.; Gosselin, D.; and Mitchell, J. L. 2004. Real-time skin rendering on graphics hardware. In *Proceedings of ACM SIGGRAPH*.

[Sun 2001] Sun, R. 2001. *Duality of the Mind*. Mahwah, NJ: Lawrence Erlbaum Associates.

[Waters 1987] Waters, K. 1987. A muscle model for animating three-dimensional facial expression. In *Proceedings of ACM SIGGRAPH*, volume 21, 17–24.

[Yang & Bringsjord 2001a] Yang, Y., and Bringsjord, S. 2001a. Mental metalogic: A new paradigm for psychology of reasoning. In *Proceedings of the Third International Conference on Cognitive Science (ICCS 2001)*. Hefei, China: Press of the University of Science and Technology of China. 199–204.

[Yang & Bringsjord 2001b] Yang, Y., and Bringsjord, S. 2001b. The mental possible worlds mechanism: A new method for analyzing logical reasoning problems on the gre. In *Proceedings of the Third International Conference on Cognitive Science (ICCS 2001)*. Hefei, China: Press of the University of Science and Technology of China. 205–210.

[Yang & Bringsjord 2005] Yang, Y., and Bringsjord, S. 2005. *Mental Metalogic: A New, Unifying Theory of Human and Machine Reasoning*. Mahway, NJ: Erlbaum.

[Yang & Johnson-Laird 2000a] Yang, Y., and Johnson-Laird, P. N. 2000a. How to eliminate illusions in quantified reasoning. *Memory and Cognition* 28(6):1050–1059.

[Yang & Johnson-Laird 2000b] Yang, Y., and Johnson-Laird, P. N. 2000b. Illusory inferences with quantified assertions. *Memory and Cognition* 28(3):452–465.

[Yang, Braine, & O'Brien 1998] Yang, Y.; Braine, M.; and O'Brien, D. 1998. Some empirical justification of one predicate-logic model. In Braine, M., and O'Brien, D., eds., *Mental Logic*. Mahwah, NJ: Lawrence Erlbaum Associates. 333–365.

EUROSIS Game-On

www.dmu.ac.uk
DE MONTFORT UNIVERSITY
LEICESTER • BEDFORD

We, the Conference Committee of the GAMEON'2005 Conference, which was held in Leicester, United Kingdom from November 24-25, 2005, hereby declare that after a peer review

Selmer Bringsjord, Sangeet Khemlani, Konstantine Arkoudas, Chris McEvoy, Marc Destefano and Matthew Daigle

were selected with their paper entitled

Advanced Synthetic Characters, Evil, and E*

As Best Paper of the aforementioned conference

Marwan Al-Akaidi	Abdennour El-Rhalibi
GAMEON'2005 General Conference Chair	GAMEON'2005 General Program Chair