The Contemporary Craft of Creating Characters Meets Today's Cognitive Architectures: A Case Study in Expressivity*

Selmer Bringsjord • John Licato • Alexander Bringsjord

version of 0121161500NY

Contents

---

[*] We are indebted to two anonymous referees, and editor Jeremy Turner, for insightful suggestions and objections. Any remaining deficiencies are due to our own failings.

Abstract

We herein investigate the following question:

(Q3) What does the contemporary craft of character design (by human authors), which is beyond the reach of foreseeable AI, and which isn't powered by any stunning, speculative, AI-infused technology (immersive or otherwise), but is instead aided by tried-and-true "AI-less" software tools and immemorial techniques that are still routinely taught today, imply with respect to today's computational cognitive architectures?

Because this is a large question, we narrow its scope so as to be able to productively consider it in the space of but the present chapter. Our narrowing is accomplished by making three moves. First, we anchor the discussion to a simple, single story (*Double-Minded Man*) populated by only a pair of characters, and crafted in a particular "AI-less" software system taught and used by scriptwriters today. Second, we focus specifically on the **expressivity** of a cognitive architecture, in light, specifically, of two expressivity challenges that arise from considering *Double-Minded Man* and its characters. And third, we partition cognitive architectures, with respect to these challenges, into three categories. The first category is composed of those architectures that appear to have no chance of meeting the challenges; the second, of those that *perhaps* meet the challenges; and the third, of those that clearly meet the challenges. As shall be seen, the cognitive architecture ACT-R falls into the second category, and CLARION falls into the third. Encapsulating, then, here's the core of our answer to (Q3):

(A_{Q3}) While ACT-R and cognitive architectures that have its level of expresivity or less are perhaps able to support cognitive models of characters in narrative, CLARION certainly has a level of expressivity sufficient to capture such characters. While we know of no other cognitive architecture that has this high level of expressivity, any architecture with this level of expressivity would at least be a candidate for the robust modeling implied by contemporary story and character design.

1 Introduction

Ibsen's characters are deep, memorable, and often dark; a contemporary Norwegian writer, Ullman, is following suit.[1] With such fiction in mind, one might ask:

(Q1) Could an AI generate robust and engaging characters like these?

(Q2) Could revolutionary AI and sensory-immersion technology augment the ability of human writers to generate robust and engaging characters like these?

Elsewhere, answers, at least provisional ones, have been offered to this pair. In the case of (Q1), S. Bringsjord (2001) has defended a negative (a view not neccesarily in concord with those of Licato); in the case of (Q2), Bringsjord & Bringsjord (2009) have defended an affirmative.[3] In the present chapter, we are interested in a very different (and — forgive us — rather longer) question, one firmly rooted in the here and now, to wit:

(Q3) What does the contemporary craft of character design (by human authors), which is beyond the reach of foreseeable AI, and which isn't powered by any stunning, speculative, AI-infused technology (immersive or otherwise), but is instead aided by tried-and-true "AI-less"[4] software tools and immemorial techniques that are still routinely taught today, imply with respect to today's computational cognitive architectures?

(Q1) is a question for AI researchers and engineers. (Q2) is a question for them as well, and for those working with them, from the field of human-computer interaction. But (Q3) isn't a question for engineers: it's a question for computational cognitive scientists, or more specifically, for computational cognitive modelers. (Q3) is relevant to those who proclaim, today, that they have on hand a "computational cognitive architecture" that captures, in one framework, most, if not all, of the nature and range of human cognition.

(Q3) can be concretized by pinning down the craft in question, and by doing the same for computational cognitive architectures and the modeling made possible by them. For the former, we accomplish this by focusing on the craft of character creation and design in the widely used, tried-and-true Movie Outline® system.[5] For the latter, we focus on two cognitive architectures: ACT-R

---

[1] In the former case, examples — such as Hakon Werle! — abound e.g. in *The Wild Duck*, readily available online; in the latter, see *The Cold Song* (Ullmann 2014), which features the endlessly fascinating Alma,

> the novels finest and thorniest creation. . . . she is by turns endearingly wistful and frighteningly willful. Love and loathing entwine in her like a double helix, and her greatest transgressions are part and parcel with her greatest strength: a fierce (possibly compulsive) insistence on telling the truth. (Cohen 2014)

These are just two examples picked from countless others. They are selected mainly because of our prior thinking about Ibsen and deep characters; e.g., see (Bringsjord 1995*b*) (and see also additional prior Bringsjordian work cited below). We recognize that some readers will be unfamiliar with the universal view that Ibsen's characters are seminally deep and robust. We regret that we don't have the space to subtantiate this view from the standpoint of literary theory and criticism, and direct readers to (Bringsjord 1995*c*) as a starting place. The fact is that what Movie Outline 3[2] requires of scriptwriters (see below) in connection with fleshing out characters is something that Ibsen did before penning a single word of a drama.

[3] There is also of course the related question

(Q3′) Could an AI generate belletristic fiction (which requires robust and engaging characters)?

A negative answer to this question is given in (Bringsjord & Ferrucci 2000); but that same work provides evidence in support of an affirmative response to a version of this question formed by supplanting 'belletristic' with 'formulaic.'

[4] By use of this adjective we mean to simply report that the kind of **intelligent agents** taken to distinguish the field of AI (Russell & Norvig 2009) are not present in the software in question.

[5] Available (including in trial form) at http://www.movieoutline.com. For the record, we have no stake in, nor do we benefit in any way financially from purchases of, this software. As mentioned above, another excellent

(Anderson 1993, Anderson & Lebiere 1998), probably the best-known cognitive architecture today, and one unmistakably aligned with the sanguine view that this architecture is well on its way to capturing the nature and breadth of human cognition (e.g. see the bold Anderson & Lebiere 2003); and on CLARION (Sun 2002), a cognitive architecture that uniquely founds its expressive power (among other things) on sub-symbolic processing, in keeping with what Sun (2001) has called "the duality of mind."

The foregoing sets our general context. Because (Q3) is a large question, we narrow its scope so as to be able to productively consider it in the space of but the present chapter. Our narrowing is accomplished by making three moves. First, we anchor the discussion to a simple, single story — *Doulbe-Minded Man* — populated by only pair of characters, and crafted in a particular software system taught and used by scriptwriters today. Second, we focus specifically on the **expressivity** of a cognitive architecture, in light, specifically, of two expressivity challenges that arise from considering *Double-Minded Man* and its characters.[6] And third, we partition cognitive architectures, with respect to these challenges, into three cateogories. The first category is composed of those architectures that appear to have no chance of meeting the two challenges; the second, of those that *perhaps* meet the challenges; and the third, of those that clearly meet the challenges. As shall be seen, the cognitive architectures ACT-R falls into the second category, and CLARION falls into the third. Summing up, then, here's the core of our answer to (Q3):

> (A_{Q3}) While ACT-R and cognitive architectures that have its level of expresivity or less are perhaps able to support cognitive models of characters in narrative, CLARION certainly has a level of expressivity sufficient to capture such characters. While we know of no other cognitive architecture that has this high level of expressivity, any architecture with this level of expressivity would at least be a candidate for the robust modeling implied by contemporary story and character design.

The basis for (A_{Q3}) is a purported demonstration that the nature of characters created in Movie Outline 3 for even a short-short film generates acute challenges. *A fortiori*, full-length works would deliver the same moral.

The structure of the sequel of the present chapter is straightforward: We first provide a brief, self-contained overview of the concept of expressivity (§2). Next, in section 3, we provide a synopsis of *Double-Minded Man*, and by doing so introduce its only two characters: Harriet and Joseph; summarize the demands made by Movie Outline 3 on the author who wishes to create rich characters; and present a pair of specific expressivity challenges arising from the composition of *Double-Minded Man* in Movie Outline 3. We then briefly discuss the category of cognitive architectures that: have no chance to meet the two expressivity challenges (§4); have perhaps a chance of meeting the two challenges (§5); can without question meet the two challenges (§6). As an exemplar of architectures in the second of these categories we use ACT/ACT-R (Anderson & Lebiere 2003), and as an exemplar of architectures in the third category we use CLARION (Sun 2003). Next (§7), we address a number of concerns and objections. We end (§8) with a brief section devoted to summing up the moral of our investigation, and pointing toward where its future lies.

---

system, which we could have used for the present paper, and for composing *Double-Minded Man*, is Dramatica ((http://dramatica.com; also available in trial form).

[6] The notion of high expressivity as discussed in the present chapter is, we cheerfully concede, ultimately a single metric for the examination of a cognitive architecture, and this metric is only one piece of a bigger puzzle — a puzzle the fuller exploration of which is beyond our scope and available space.

2 What is Expressivity?

Perhaps the best way to approach the topic of expressivity is to treat English declarative sentences as input to an algorithm $\mathcal{R}$ that outputs the formalized *meaning* of the input sentences, in the form of representations of those sentences, where these representations are expressed as formulae in the formal language of some logic placed within some received continuum.[7] For instance, there is the well-known continuum that begins with zero-order logic (ZOL), then moves to first-order logic (FOL), then to second-order logic (SOL), then to third-order (TOL), and so on; each of these is more expressive than its predecessors[8] We can then view $\mathcal{R}$ as computing a function, and accordingly write $\mathcal{R} : \sigma \mapsto \phi_{\sigma\mathcal{L}}$, where σ is an English declarative sentence, and $\phi_{\sigma\mathcal{L}}$ is a formula in the particular logic $\mathcal{L}$ that's in play. Even for readers unfamiliar with the continuum just alluded to, it's easy to get the hang of it in general, and easy to get the hang of how it works in connection with $\mathcal{R}$. Here's how.

Consider first the following sentence, which is true in *Double-Minded Man*.

(A) Harriet loves Joseph.

We can easily represent this sentence in ZOL.[9] ZOL allows us to use **relation symbols** to represent relations (or properties), **constants** to represent individual objects, and to then build formulae that ascribe these relation symbols to these constants. This logic also includes the familiar quintet of **truth-functional connectives**: $\wedge$ ('and'), $\vee$ ('or'), $\rightarrow$ ('if _ then _'), and $\leftrightarrow$ ('_ if and only if _'). No quantification is available in ZOL, and there are no variables (but identity, =, is included — though we don't use it herein; with = comes inference schemas for reasoning with identity). For handling (A), we need a relation symbol to represent the two-place property of one thing loving another, and we need two constants, one to denote Harriet and one to denote Joseph. With these ingredients and some straightforward punctuation, we can specify $\mathcal{R}[(A)]$:

(A′) $L(h, j)$

Now here's an example of a more expressive English sentence that is also true with respect to the story *Double-Minded Man*:

(B) Joseph owns a BMW sportbike.

To represent (B), we can use the following formula in FOL:[10]

---

[7] The algorithm $\mathcal{R}$ can be applied to any declarative sentence, irrespective of the role that that sentence might have. In cognitive models arising from processing in cognitive architectures, some information in declarative memory might be believed or acted on without any deeper reasoning or decision-making. Algorithm $\mathcal{R}$ takes no account of such shades. The same would go for the practice of mathematics: Some declarative sentences involved in the practice of mathematics are largely ignored; others receive a lot of attention. Yet $\mathcal{R}$ can be applied to 1+1=2 just the same as to some great unsolved conjecture, such as that $\mathbf{P} \neq \mathbf{NP}$.

[8] The continuum is e.g. presented cogently in (Andrews 2002).

[9] ZOL subsumes the propositional calculus.

[10] Cognoscenti in pursuit of building $\mathcal{R}$ would call for pairing FOL with e.g. the additional machinery of the λ-calculus, as e.g. in (Blackburn & Bos 2005). But we can leave this aside. Readers interested in the formal tools in linguistics needed for intelligent pursuit of $\mathcal{R}$ should begin their study with (Partee, Meulen & Wall 1990). By our lights, such material should be in every toolbox of every cognitive modeler. After all, the human ability to understand the meaning of natural language is a key part of what it is to be us, cognitively speaking.

(B′) $\exists x(B(x) \land S(x) \land O(j,x))$

The symbol '$\exists$' is the well-known **existential quantifier**, and when paired with variable x, as in '$\exists x$,' can be read as 'there exists an x such that.' As to B, S, and O in (A′), they of course are relation symbols that represent, resp., the properies *being BMW-manufactured*, *being a sportbike*, and *being a thing _ that owns a thing _*. FOL, as most readers will doubtless recall, also incudes the **universal quantifier**, $\forall$, which allows us to capture such sentences as

(C) Every spouse of Joseph owns more than one Bible.

which is also true of Harriet in *Double-Minded Man*,[11] and would by $\mathcal{R}$ yield

(C′) $\forall x[Sp(x,j) \rightarrow (\exists x \exists y \exists z(Bi(x) \land Bi(y) \land O(z,x) \land O(z,y)))]$

To be clear about the aforementioned continuum: We invoked ZOL by regimenting the ascription of properties to particular individual things picked out by constants, then moved from ZOL to FOL by allowing quantification over lower-case variables $x, y, \ldots$ that are placeholders for individual objects. To next move from FOL to SOL, we simply extend quantification so that it ranges over not only individual objects, but over relation symbols as well. To do this, we simply introduce a new collection of majuscule variables — $X, Y, Z, \ldots$ — that range over not individual objects, but over relation symbols (and hence ultimately over properties). To quickly get a sense of what this specifically buys us, consider an English sentence that communicates another truth about *Double-Minded Man*:

(D) There are attributes Joseph has that Harriet's mental model of him doesn't.

Let's denote Harriet's mental model of Joseph by m. Then we can represent (D) in SOL as follows:

(D′) $\exists X(X(j) \land \neg X(m))$

A move, at this point, to TOL would consist in a direct analogue of what got us started in the first place when we built ZOL: viz., we allow ascription of a new class of relation symbols to our original set of relation symbols. So far, relation symbols have been pulled from the set of upper-case Roman letters, and applied to individual objects only. For our new class of relation symbols, which allow us to ascribe properties to properties, we can avail ourselves of $\mathcal{F}, \mathcal{R}, \mathcal{S}, \ldots$; we put these symbols to use below, when issuing the first of our two promised expressivity challenges.

The second of those challenges requires the deployment of machinery from a different class of logics: **epistemic** logics.[12] Whereas the FOL-to-TOL sequence, as well as the logics that follow in the continuum, are well-suited to modeling non-cognitive domains like number theory, epistemic logics are marked by the addition of an entirely new device: **operators**. The two key operators in epistemic logic are **K** and **B**, which represent, resp., 'knows' and 'believes.' In order to add these operators to our ZOL-to-TOL sequence systematically, we have only to affirm the following grammatical rule:

---

[11] This fact isn't *revealed* in *DMM*, but is something which, in keeping with the authorial craft of creating robust characters, is in the author's mind when writing. For more on such techniques and Ibsen, see (Bringsjord 1995*c*).

[12] Coverage of such logics in connection with cognitive modeling is provided in (Bringsjord 2008). For an introduction to epistemic logic and other logics with operators see (Goble 2001).

(R) If ϕ is a well-formed formula, then so are $\mathbf{B}_c\phi$ and $\mathbf{K}_c\phi$, where c is a constant denoting a member of a reserved set of constant symbols used to denote agents.

Note that operators can be iterated. In (Bringsjord & Ferrucci 2000), an example from rudimentary mystery fiction is used to show that, where d denotes the detective, v the villain the detective is seeking to find and bring to justice, and V is a relation symbol representing the property *being a villain*, a formula like the following one represents something that writers of such fiction routinely know in the course of their writing:

(V) $\mathbf{B}_d\mathbf{B}_v\mathbf{B}_d V(v)$

3 *Double-Minded Man*, and Expressivity Challenges Therefrom

3.1 Gist of *Double-Minded Man*

Double-Minded Man is a two-character, three-scene short-short film script written by S. Bringsjord and A. Bringsjord in and through *Movie Outline 3*, in a way that follows the approach to creative writing described in (Bringsjord 1992). The term ‘Double-Minded’ in the title refers to the mind of Joseph, by all accounts for years a rather boring and morally upright financial advisor working out of his own small suburban firm, but who is in reality a hard-drinking, cocaine-and-heroin-using, sportbike-driving, decidedly non-monogamous chap. Here’s the current version of the first scene:

```
TWIRL - DAY

68-year-old Harriet Smith sits with two wrinkled hands firmly on the
wheel of her rust-eaten Subaru wagon, staring straight ahead through
the top level of bifocals as she waits serenely at a red light.

Harriet is alone in the car. To her right is another vehicle, also
waiting, in this case to make a right turn; it’s a sleek, low-slung,
black Camaro.

We are inside the cabin with Harriet.  The Subaru’s sound system
softly plays choral music.  Harriet’s lips move slightly as she
internally sings along, mouthing a slow aria.  Her head weaves
slightly side to side, in the rhythm with the music.

Things are calm as can be here inside the car with Harriet.  There are
a pair of well-worn Bibles on the empty passenger seat beside her,
one with a gold-lettered ’Harriet’ on its leather front cover, the
other with a matching ’Joseph’ on its front cover.

Harriet’s eyes swivel up to the light: still red.  We wait with her.

Suddenly there is a piercing SCREECH outside.  Harriet jerks her head
to the right and we follow her line of sight.

A sleek motorcycle has swerved out of its lane and is now streaking
straight for the right side of the Camaro beside Harriet’s car.
```

```
The bike slams with CLANG into the side of the Camaro.  Its rider is
flung up and forward into the air, twirling passed Harriet's
windshield.

We now watch from Harriet's POV, in slow motion.  The
black-leather-clad motorcyclist sails by Harriet's windshield,
airborne.  We see a man's face, clearly: His elephant-hide skin tells
us that he is well beyond middle-age.  Yet thick, black curls of
youthful hair emerge from under his helmet.  The rider has only one
half of a black, bushy, swept-out, waxed mustache.  His eyes are
weary and grey, and appear to lock with Harriet's for an instant.

We return to normal speed.  The body is now lying on the incoming lane
to the left of Harriet's Subaru, perfectly still on the blacktop, the
head twisted into an impossible angle.  Blood seeps from a nostril.
Beside the lifeless head, a BMW medallion lies on the pavement,
glinting in the sunlight.
```

Joseph, Harriet's 69-year-old husband, is thus dead on the spot. As we learn in the next scene, set in a morgue, Joseph's real hair consists of but anemic, white fuzz, and the second half of his fake mustache has been removed. The third and final scene is Harriet's visit to Joseph's secret lair, a small and dingy storage unit where he parked his sleek and super-fast motorcycle in secret. She there comes upon a number of things that reveal the dark aspects of Joseph's secret life, of which, hitherto, she had not the slightest inkling.

3.2 Basic Demands of Movie Outline 3

Movie Outline 3, like its predecessors 1 and 2, is a system that makes instructively serious demands of the creative writer who would use it. This is especially true with respect to the *characters* in a story created and crafted by the creative writer. For example, the creation of a character requires that 16 initial data-fields be completed. One of these fields is `Education`, and in Joseph's case it's recorded that he has a BS in Math, an MBA in Finance, and is a Chartered Financial Planner (CFP); but none of these details appear in the script itself. Notice that we here say *initial* data: Movie Outline 3 asks for no less thatn 68 deep questions in order to evince and record rich information about a given character; this is the so-called `Interview` component of character construction in the system. There are other aspects of character design encouraged by Movie Outline 3, but what we say here is sufficient, we trust, to allow our readers to understand that the pair of expressivity challenges to which we now turn are not in the least exotic. It would be easy, and fair given what Movie Outline 3 requests, to present much more demanding expressivity challenges. Parallel propositions would arise in connection with the creation in Movie Outline 3 of *any* significant characters in *any* substantive narrative.[13]

3.3 Harriet, Joseph, and a Pair of Expressivity Challenges

In *Double-Minded Man*, the following sentence holds:

---

[13] We offer here no formal definition of 'substantive narrative.' There is a longstanding tradition in AI to count such things as 'Bill went to the store and bought a candy bar' as a story (e.g. see Charniak & McDermott 1985). But clearly such a thing isn't a *substantive* story.

(1) Joseph knows at some time before his death that his wife Harriet doesn't know (at that same time) that Joseph has a racy lair.

At a minimum, when $\mathcal{R}$ is applied to (1), we obtain:

(1′) $\mathbf{K}_j(t < d \land \neg \mathbf{K}_h(t, R(j)))$

As to TOL, there is the following, necessarily known by the authors of *DMM*:

(2) There are attributes Joseph has that are dark and racy.

For (2), notice that we need to quantify over relations (SOL), and we reach TOL because one or more of these relations has the property of being dark-and-racy:

(2′) $\exists X(X(j) \land \mathcal{DR}(X)$

These, then, are the two expressivity challenges we had promised to issue.

4 The "No Way" Category and the Expressivity Challenges

Some cognitive architectures, as far as anyone can honestly tell, clearly are incapable of meeting expressivity challenges like (1) and (2). Since it's likely to be somewhat distressing for proponents and developers of the cognitive architectures in this category to learn that the architectures they presumably think highly of are in fact (in this regard) lowly, we tread very lightly by saying very little about membership in this category. Unfortunately, it seems to us that a number of cognitive architectures still under active development fall into the "no way" category. To give at least one example, we note that RCS (Albus 2000), which to our knowledge is not under active development, appears to be unable to represent (1)/(1′) and (2)/(2′). The reason is simple. The parts of RCS designed to hold representations of the external world in a knowledge-base, to use terms Albus (2000) employs, the parts of RCS that include "frames, objects, classes, and rules," are not associated with any such machinery as complex quantification, application of properties to properties, and operators added to the mix as well.

5 The "Maybe" Catgory ACT-R, and the Expressivity Challenges

Turning now to ACT-R, even under the most charitable assumptions, it appears incapable of expressing (1′) and (2′), and hence appears incapable of representing the cognition that arises from writing or reading and understanding *Double-Minded Man*. This is so for the simple reason that at most ACT-R (as a descendant of ACT) has the expressivity of FOL, as shown in (Anderson 1976). We say 'at most' because for example the process described by Anderson (1976) for converting a formula ϕ_σ in FOL to an expression in ACT passes through a phase where $\phi_{\sigma\mathcal{L}}$ (where $\mathcal{L}$ = FOL) is replaced by a corrresponding formula $\phi^{\text{SNF}}_{\sigma\mathcal{L}}$ in **Skolem Normal Form**. But it's a well-known theorem that the meaning of $\phi^{\text{SNF}}_{\sigma\mathcal{L}}$ doesn't in the arbitrary case correspond to the meaning of $\phi_{\sigma\mathcal{L}}$. That is, it's not the case that any model $\mathcal{M}$ satisfying $\phi_{\sigma\mathcal{L}}$ also satisfies $\phi^{\text{SNF}}_{\sigma\mathcal{L}}$, and *vice versa*.

Other work devoted to providing a "formal semantics" for ACT-R since 1976 has not covered expressivity in our sense, and has nothing to do with formal semantics as operationalized in the logician's and linguist's sense of $\mathcal{R}$. For example, Gall & Frühwirth (2014) give, as they explain, merely an *operational* formal semantics for ACT-R. Theirs is an investigation of the computational processes associated with production-system-processing, not of the expressivity of ACT/ACT-R with respect to formulae as the formal bearers of the meaning of declarative sentences in natural languages like English.

Note that it's very hard to see how identity in the semantics provided for ACT in (Anderson 1976) provides the power of FOL. For example, the first prime number greater than 3 is identical to the sum of 4 and 1. This fact is easily captured in FOL by for example

$$\exists x[P(x) \land x > 3 \land \neg\exists y(P(y) \land y < x) \land x = +(4,1)]$$

It doesn't seem possible for simple arithmetical facts like this to be represented directly in ACT/ACT-R, in such a way that meaning is preserved, even leaving aside the apparently fatal problem we have already cited. This is so because the way identity works in FOL doesn't seem to be carried over to ACT/ACT-R. In FOL, for that matter in ZOL, if we have that $R(a)$ and that $a = b$, the upshot is that exactly the same models render $R(b)$ true as render $R(a)$ true.[14] Yet Anderson (1976) writes (pp. 224–225) that in ACT, even given that $a = b$, $R(a)$ and $R(b)$ don't have the same meaning. Yet these formula *do* have the same mearning in ZOL and FOL.

Furthermore, the sense of expressivity used by the present paper should be distinguished from the idea of *knowledge level* (Newell 1981). Soar, for example, can justifiably boast of its ability to represent and reason over a wide variety of types of knowledge that vary qualitatively: procedural, episodic, declarative, and so on (Rosenbloom, Newell & Laird 2014). But like ACT/ACT-R, little to no work has been done in demonstrating that these representations have a logical expressivity equivalent to FOL, much less second- or any higher-order logics. In short, then, the metric that we are concerned to apply and discuss herein is one that hasn't been addressed. We don't know how Soar would measure on this metric, but one can of course view the present chapter as, in the end, a call for a wider application of the rigorous metric that is our concern. Put another way, since our metric is standard fare in measuring the relative expressivity of precise ways of expressing declarative information (e.g. see Lindström's Theorems in Ebbinghaus, Flum & Thomas 1984), it would appear that the Soar communinity hasn't yet measured the capacity of Soar to represent declarative information.

Finally, note that while (as we have just noted) Anderson (1976) discusses expressions having `believes` within them, this string is treated no differently than any other predicate, and the referential opacity of belief in English isn't formalized.[15] There would be two ways to carry out such formalization. The first way would be through some machinery tailor-made for the purpose, and beyond standard model theory (or a variant) for ordinary first-order logic; for instance, through a possible-worlds semantics (which is often used for epistemic logic; e.g., see (Fagin, Halpern, Moses & Vardi 2004)). But we read:

> Development of a possible worlds semantics may also be a better way to deal with the opacity of

---

[14] See any standard textbook for confirmation. E.g., (Barwise & Etchemendy 1999).

[15] In fact, the idiosyncratic treatment of identify that we noted in the previous paragraph is declared to be necessary to block such inferences as that if someone believes $R(a)$, and $a = b$, that same someone would then have to believe $R(b)$. (Given that Smith believes that John Le Carré is the wealthy author of *The Spy Who Came in From the Cold*, and that JL = David Corwell (which is true), it would follow that Smith believes DC to be wealthy. But Smith may not know that 'Le Carré' is a pen name.) But this unwanted consequence only arrives if belief is formalized naïvely, instead of in an operator-based logic like epistemic logic.

> propositions just discussed. However, I do not feel technically competent to develop a possible worlds semantics. (Anderson 1976, p. 225)

Neither is an alternative pursued. For instance, one might pursue a proof-theoretic semantics (see e.g. Bringsjord, Govindarajulu, Ellis, McCarty & Licato 2014), but that isn't done by Anderson (1976).

Despite the foregoing problems, we rest content out of charity to make only the circumspect claim that it certainly *seems* to be the case that ACT-R cannot meet our two expressivity challenges. After all, the thrust of (Anderson 1976) is that ACT has the expressivity of FOL, not TOL, and certainly not TOL in conjunction with operators.

We turn now to CLARION, and it's capacity to meet our pair of challenges.

6 CLARION and the Expressivity Challenges

6.1 Overview of CLARION

CLARION (Sun 2002) is an integrative cognitive architecture that has a dual-process structure consisting of two levels: explicit (top level) and implicit (bottom level). These levels roughly correspond to the localist-distributed split we described earlier. CLARION has been able to model a wide variety of cognitive phenomena while maintaining psychologically plausible data structures and algorithms; this makes it an ideal choice for our purposes. By showing that structured reasoning can emerge from no more than the mechanisms in CLARION which previous literature have already shown to be psychologically plausible, we provide a strong foundation for showing that these new structures are psychologically plausible as well.

The architecture is further divided into four *subsystems*, each with explicit and implicit levels, which specialize in different aspects of cognition: The Motivational Subsystem (MS), the Metacognitive Subsystem (MCS), the Action-Centered Subsystem (ACS), and the Non-Action-Centered Subsystem (NACS). We will be focusing on the NACS in the present essay.

6.1.1 NACS — the Non-Action-Centered Subsystem

The NACS contains general knowledge about the world that is not contained in the ACS. Whereas the ACS is meant to capture the knowledge that directly causes decision-making while interacting with the world, the knowledge in the NACS is often more deliberative and is used for making inferences. The top level of the NACS is called the General Knowledge Store (GKS), and it contains localist chunks that can be linked to each other using Associative Rules (ARs).

The bottom level of the NACS is called the AMN, or the Associative Memory Network, and it contains implicit associative knowledge encoded as dimension-value pairs (DV pairs). Each GKS chunk is connected to a set of DV pairs in the AMN with some weight that can be adjusted over time. This unique structure gives CLARION the ability to define a *directed* similarity measure between two chunks c_1 and c_2, which is derived from the amount of overlap between the DV pairs connected to the two chunks (Sun 1995, Tversky 1977, Sun & Zhang 2004):

$$S_{c_1 \to c_2} = \frac{\sum_{i \in c_2 \cap c_1} W_i^{c_2} \times A_i}{f(\sum_{i \in c_2} W_i^{c_2} \times A_i)} \tag{1}$$

where $f(x) = x^{1.0001}$. Sun and Zhang (2004) define A_i as the strength of activation of the values of dimension i in chunk c_1, and $W_i^{c_2}$ as the weights of the DV pairs specified with respect to c_2. However, in this chapter we simplify things by setting all A and W values to 1, which reduces Equation 1 to a function of the number of dv pairs connected to c_1 and c_2:

$$S_{c_1 \to c_2} = \frac{|c_1 \cap c_2|}{|c_2|^{1.0001}} \tag{2}$$

Note that it is possible for the denominator in Equation 2 to be zero, in which case the entire equation is given the default value of 1.

The Associative Rules (ARs) link groups of chunks to other chunks in the GKS, and consist of a set of condition chunks $c_1, c_2, ...$ and a single conclusion chunk d. For any given AR, each condition chunk i has a weight W_i such that $\sum_i W_i = 1$. We write out a single associative rule in the following format:

$$(c_1, c_2, ..., c_n) \Rightarrow d$$

The chunks in the GKS and DV pairs in the AMN have activation levels which can be set by CLARION's other subsystems. Activations can also spread through the NACS using the chunk-DV pair connections and the top-level ARs. The manner in which this activation spreads can be restricted: other subsystems can temporarily disable Rule-Based Reasoning (activation spreading through ARs) or Similarity-Based Reasoning (activation spreading through chunk similarity), or perform activation propagation as some weighted combination of both of these reasoning types. These abilities are detailed further in Sun & Zhang (2004, 2006), in which these mechanisms are shown to be psychologically plausible by using them to closely emulate the results of psychological studies. We use no more than these mechanisms to construct the knowledge structures in this chapter.

6.2 The Expressivity of CLARION

6.2.1 Representing Structured Knowledge in General

The associative rules linking NACS chunks already seem to impart a kind of weak structure to the GKS, but they do not constitute structural knowledge in the sense described by, for example, (Hummel & Holyoak 1997, Holyoak & Hummel 2000, Licato 2015). In order to impart structure, and to allow CLARION to represent knowledge with high levels of expressivity, (Licato, Sun & Bringsjord 2014*a*) introduced *types* of chunks. The scope of these types, however, holds only within the context of a complete structure, for reasons we will explain shortly. It is convenient to start by basing our structures on the well-established S-expressions, which are perhaps most notable for their use in the programming language LISP. An S-expression is of the form:

$$(P \;\; o_1 \;\; o_2 \;\; ... \;\; o_n)$$

where P is a predicate, and each o_i is either an object or another S-expression. Our first chunk type, then, is the object chunk. We also define a proposition chunk, which is both a marker of the relationship between the object chunks, and a placeholder for the proposition's predicate symbol. The proposition and object chunks are pictured in Figure 1 as oval-shaped objects.

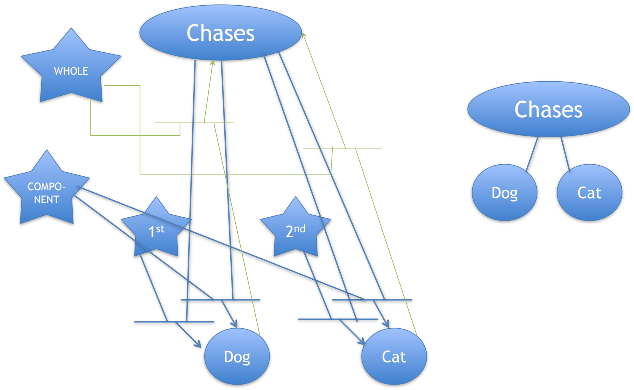

Figure 1: A knowledge structure representing the proposition $CHASES(DOG, CAT)$. On the right is the simplified version, which omits the CDCs and many of the ARs, though they are there (just not pictured). (Things are kept generic here via reference to the proverbial dog-cat scenario. Building structures specifically tailored to *Double-Minded Man* we judge to be otiose in the present context — but see our return to this issue in the final section of the present chapter.)

Of course, something needs to link these chunks together, and that is where Cognitively Distinguished Chunks (CDCs) come in. Given that all neurobiologically normal adult humans are capable of performing structured reasoning, we should assume that there are some common cognitive abilities that are either innate or develop very early in life which allow for structured knowledge to emerge. CDCs are meant to reflect these abilities, and we maintain psychological plausibility by placing the following restrictions on them. Firstly, CDCs are fixed: We do not define any algorithms that create or destroy CDCs. Secondly, CDCs are known to basic reasoning algorithms. For example, the algorithm we describe later in this chapter which performs analogical reasoning can refer to certain CDCs directly, under the assumption that these are basic features of structured knowledge. Finally, if there is a function that can be easily performed using a CDC, then that function is assumed to be a basic ability of any neurobiologically normal adult human reasoner. This final point will be elaborated shortly.

CDCs are depicted as star-shaped (Figure 1). Associative rules link the CDCs to the chunks in the structure. For example, the *WHOLE* CDC links object nodes to proposition nodes. In Figure 1, which depicts the proposition *CHASES(DOG CAT)*, the *WHOLE* CDC is part of two ARs (depicted in the Figure as an arrow with multiple tails and one head):

$$(DOG, WHOLE) \Rightarrow CHASES$$
$$(CAT, WHOLE) \Rightarrow CHASES$$

Each object chunk in the condition of these ARs has a weight of 0.5. In fact, for all ARs we mention in the present chapter, the weight is distributed evenly amongst all objects in the AR's condition unless otherwise mentioned. If an AR contains a CDC as one of its condition chunks, then we call that AR a *CDC-based AR*. As we mentioned earlier, one limitation on CDCs is that CDC-based ARs should correspond to basic abilities of neurobiologically normal adult human reasoners. This means that given some chunks in a knowledge structure, if it is easy to traverse CDC-based ARs and retrieve some other part of that structure, then that retrieval should also be something humans can do easily. The "basic ability" that the above two ARs correspond to is the ability

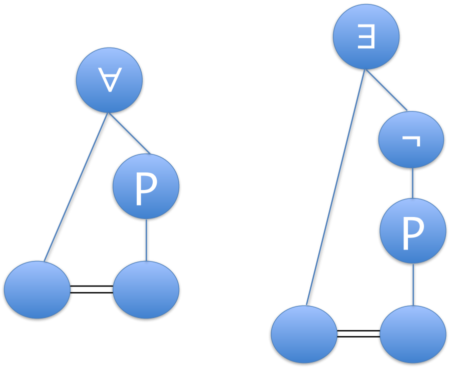

Figure 2: First-order-logic formulas $\forall_x P(x)$ and $\exists_x \neg P(x)$.

to recall propositions involving an object, given nothing but that object. Imagine being asked to recall some fact about dogs. Among others, one of the facts that likely would be recalled is that dogs chase cats (assuming, of course, that the reasoner in question is aware of this fact). That is modeled here by activating the *WHOLE* and *DOG* chunks. The activation would spread through any ARs which contain those two chunks in their conditions, and the resulting proposition nodes would be activated.

A *COMPONENT* CDC is also defined to introduce some redundancy into the structure, such that for every rule involving a *WHOLE* CDC, a complementary rule going in the other direction is created with a *COMPONENT* CDC. Whole chunks are always pictured above component chunks. To preserve argument order, we introduce *Ordinal CDCs*, which are also pictured in Figure 1 as *1ST, 2ND,* etc. Ordinal CDCs simply preserve the roles objects play within propositions in a general way that does not name the roles specifically (contrast this with the LISA model (Hummel & Holyoak 2003), which has distinct role units for every type of role).

The basic proposition structure we have been describing can also be nested, so that instead of an object chunk a proposition chunk can have another proposition chunk as a component. A proposition chunk can even have a single object chunk as a component multiple times, as would be necessary in the proposition $P(a, X, a)$ (Figure 4). We now move on to describing how reasoning can be performed over these structures.

6.2.2 FOL-level Expressivity in CLARION

The work described in this chapter was designed in part to be at least as expressive as first-order logic (FOL), the touchstone for assessing the expressivity of extensional logics (Ebbinghaus, Flum & Thomas 1994).[16] In this section we show that a major part of the goal to at least reach FOL has been met. We do this by first showing how the full syntax of FOL can be represented in our knowledge structures.[17]

---

[16] Extensional logics do not permit non-extensional operators like *believes*, which e.g. can be applied to propositions ϕ whose semantic values do not predictably generate a semantic value for the "outer" proposition. This is why, earlier, we stepped outside the ZOL-to-TOL sequence to epistemic logic. E.g., even if we know that ϕ is false, we cannot infer that 'Jones believes ϕ' is false (or, for that matter, that it's true). For introductory discussion of this phenomenon in the context of computational cognitive modeling, see (Bringsjord 2008).

[17] We spare the reader a straightforward proof by induction over terms and formulae in first-order logic, in which one demonstrates that every formula in a countably infinite progression of all first-order formulae can be represented

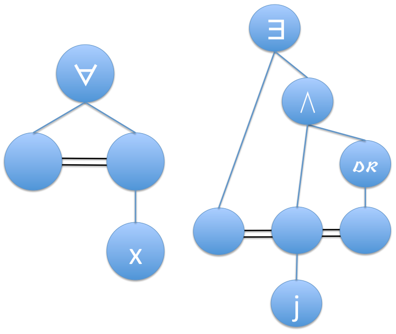

Figure 3: The proposed representations of formulae in SOL (left) and TOL (right) are pictured above. These particular structures correspond to the formulae $\forall_P P(x)$ and $\exists_X(X(j) \land \mathcal{DR}(X))$

In order to represent FOL formulae, we adopt a structure that directly maps to human-readable syntax. We represent the universal and existential quantifiers as if they were higher-level predicates, using identity links to connect quantified variables to instantiations within the variable's range. This can be seen in Figure 2, along with an example of negation, which is similarly treated as a kind of single-place predicate.

Such structures can exist in this format for easy recall by the reasoner. When they are to be used in active reasoning processes, however, the structures can be recalled from long-term memory and transformed into a form more amenable to reasoning processes (e.g., to the structures in Figure 2), perhaps by some procedure which originates in one of CLARION's other subsystems. Such a process is a bit more involved than the examples we demonstrate here. For example, this process needs to distinguish between the chunks corresponding to the universal and existential quantifiers in order to treat their corresponding structures differently in inferences.

Allowing for a native representation that contains quantifiers is a first step in simulating a so-called "natural" reasoning process, that is, a set of mechanisms that are known to better correspond to how humans, as opposed to machines, reason (e.g., something akin to *natural deduction*, introduced in (Jaśkowski 1934)). Aiming at natural reasoning may seem an odd choice, considering that modern automated theorem provers tend to prefer methods such as first-order resolution, but we remind the reader that our goal here is to model reasoning in a psychologically plausible way — and in a way that integrates with sub-symbolic processing in the human system.

6.2.3 Challenges (1) and (2) Met

By expanding on the structuring method used to represent FOL formulae in the NACS, we can represent formulae in SOL and TOL (Figure 3). For SOL, the expansion requires that the first argument under the quantifier chunk (pictured as the leftmost child of the chunk corresponding to the ∀ quantifier) is connected, via an identity link, to another blank chunk that serves the role of a predicate rather than an object. A similar change makes TOL representation possible.

It should be noted that the difficultly in ensuring cognitive architectures have higher-level expressivity is not in manually creating tree-like structures that resemble formulae in higher-order logics. That can be done rather trivially. Rather, giving a cognitive architecture the expressivity of a higher-order logic is about accurately capturing the *inferential abilities* of that higher-order logic:

---

in the scheme we have introduced.

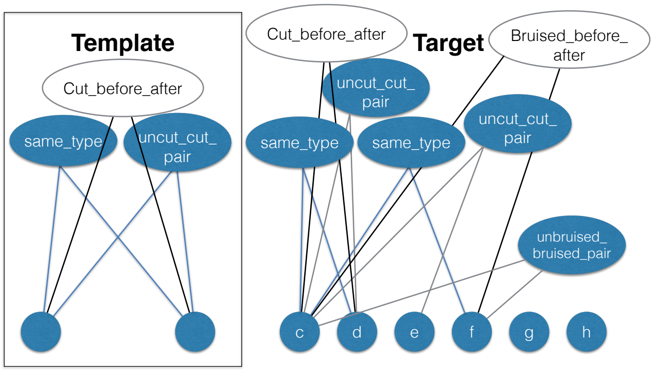

Figure 4: An example of the template structure allowing inferences in CLARION's NACS.

being able to carry out inferences or produce new structures, if *and only if* the formal semantics of the logic allows them.[18]

Ensuring that an inferential system only carries out inferences which it should be allowed to requires, at minimum, an ability to very carefully reason over deep structural properties of a representation. CLARION's NACS — again, due to work in (Licato et al. 2014*a*) and (Licato, Sun & Bringsjord 2014*b*) — makes such controlled structure-based inference possible, by making use of what are called *templates*, pictured in Figure 4. A template is simply an organization of NACS structures specifying what properties of a structure should hold, and if those properties hold, how inferences should be created. (Licato et al. 2014*a*) used templates to produce basic deductive and analogical inferences, and (Licato, Sun & Bringsjord 2014*c*) made use of templates to create explanations by generating chains of inferences.

7 Concerns & Objections; Replies

In this section we anticipate and address a series of concerns and objections.

7.1 Concern 1: "What is the purpose?"

The first anticipated concern is as follows: "Its not entirely clear what the intended target for the virtual characters in the work is. Is it for social simulation? Interactive fiction? Video games? It does make a difference in this case, for after all, a cognitive architecture needn't support all of the logic models and devices you discuss in in order for that architecture to be suitable for certain purposes. For example, if Im watching a story being acted out by AI, I dont really care what amount of cognition is going on or is supported by the underlying system; the only thing that matters is the quality of the performance. Do I believe in the characters and their circumstances? Am I engaged, emotionally, in the proceedings, and so on? How the system in the end works is secondary in this kind of application. In other applications, what is going on behind the scenes can be more important."

We reply as follows.

---

[18] Recall our earlier remarks about a failure of deduction using identity in ACT-R.

The concern as articulated here appears to ignore the explicitly stated purpose of our investigation, which is to seek to provide a substantive part of an answer to (Q3). Put briefly, (Q3) asks what the craft of character creation and design, when carriedd out in systems dedicated to framing and facilitating such creation and design, implies with respect to cognitive architectures and the science thereof. We are not concerned with the possibility of creating, designing, and deploying (in engaging narrative) characters in ways that entirely obviate the programmatic nature of the science of cognitive architectures.[19] Rather, as (Q3) makes concrete and specific, we are concerned to see how the scientific program of seeking an adequate cognitive architecture, that is, a cognitive architecture that can enable the computational modeling of human cognitive activity, fares when measured against the cognitive activity of creating, designing, and deploying characters (via sophisticated tools like Movie Outline 3).

It's important to realize that our interest in (Q3) is a particular instance of a more general question. (Q3) has a any number of close relatives, for domains other than character design and narrative. For instance, there's a version of (Q3) for the craft or creating not narrative, but introductory (differential and integral) calculus books; viz.,

(Q3)$_{calc}$ What does the contemporary craft of designing (by human authors) introductory (differential and integral) textbooks, an activity which is beyond the reach of foreseeable AI, and which isn't powered by any stunning, speculative, AI-infused technology (immersive or otherwise), but is instead aided by tried-and-true "AI-less" software tools and immemorial techniques that are still routinely taught today, imply with respect to today's computational cognitive architectures?

In the case of (Q3)$_{calc}$, as in the case of (Q3), expressivity is a key issue. One reason is that a central concept that must be understood by any human engaging in the craft of designing the textbooks in question is that of a **limit**, defined in Bolzano-Cauchy-Weierstrass (B-C-W) style. Confirmation of this route is easy to obtain, by empirical means. For example, in the 10th edition of the best-selling *Calculus* (Larson & Edwards 2014) we read, at the outset:

> So, one way to answer the question "What is calculus?" is to say that calculus is a "limit machine" that involves three stages. The first stage is precalculus mathematics, such as the slope of a line or the area of a rectangle. The second stage is the limit process, and the third stage is a new calculus formulation, such as a derivative or integral. (Larson & Edwards 2014, p. 42)

The very same conception can be found in every single other widely used calculus textbook.[20] Yet it is a logico-mathematical fact that writing the definition of a B-C-W limit requires writing out (and presumably understanding) extremely robust levels of quantification. In fact, even basic arithmetic at the level of what is learned in elementary and grade school requires robust quantification.[21]

---

[19] In light of fundamental formal facts, we cheerfully concede that there are any number of ways to create, design, and deploy characters that completely dodge the expressivity challenge, and for that matter completely dodge *all* such challenges. For example, the random generation of characters and or pixels could produce compelling narrative, replete with robust three dimensional characters. As explained elsewhere (Bringsjord 1995*a*), for any empirical, finite stimuli that a human viewer would be impressed with, there is a completely serendipitous path to producing that stimuli.

[20] The proviso 'widely used,' note, is crucial. This is so because e.g. a teacher of calculus determined to teach via infinitesimals can of course either use his/her own content, or use some published material. In the latter case, some excellent Leibnizian-Robinsonian options are available: (Keisler 1986, Sullivan 1974).

[21] The Peano Axioms for arithmetic specifically require either an infinite number of quantifier-rich formulae at the

7.2 Concern 2: “What exactly are these ‘challenges’?”

The second concern we consider is related to the first, but is more specific; it can be expressed as follows:

“Your discussion introduces two ‘challenges’ for cognitive architectures to pass, but the terms and conditions of the challenges are never made clear. For example, just because a system does not contain complex quantification, this does not mean it is not capable of supporting or meeting the challenges given in the chapter. For instance, a system does not have to be able to fully express ‘There are attributes Joseph has that are dark and racy’ for it to assign attributes to Joseph that are dark and racy, and to induce Joseph to act in that fashion. If the challenge is to have characters behave according to some kind of rich character definition, more systems than just CLARION will be able to rise to that challenge. If supporting complex quantification is in itself the challenge, then of course many systems would not meet that challenge, but then we would need to debate the relevancy or necessity of such a challenge in the first place.”

Given our reply to the previous concern, our response to the second concern is straightforward: On the contrary, the two challenges have been specified, and they cannot be dodged by any cognitive-architecture aspiring to complete the scientific program of modeling human cognition that is part and parcel of creating, designing, and deploying characters, at the level of what is requested by the likes of Movie Outline, and such as are seen in the simple *Double-Minded Man* (let alone in the dramas of Ibsen and the modern novels of Ullmann).[22] An exactly parallel point, as we have noted, can be asserted with respect to the cognition that is part and parcel of creating not fiction, but technical non-fiction (regarding calculus and arithmetic).

7.3 Concern 3: “Can CLARION handles *Double-Minded Man*?”

Here’s the third anticipated concern: “Your chapter reasonably establishes that CLARION can express the various logic systems discussed earlier. But what does that show exactly? That earlier discussion of yours is surly not the same as using CLARION to create an enactment of *Double-Minded Man*, with characters acting independently and autonomously according to their natures. And isnt that more important?”

This is a cogent concern. In keeping with our avowed purpose and plan, we rest content with having only demonstrated that CLARION *can* be used to represent the characters in *Double-Minded Man*, and that CLARION can thereby support an enactment. We haven’t built the specific representations and carried out the enactment. We have shown that this additional work can be carried out, given more time and space.

7.4 Concern 4: “What about the character-centric AI of Mateas et al.?”

The fourth concern we anticipate is as follows: “Doubtless it’s unintentional on your parts, but the foregoing will give some readers the distinct impression that the connection you forge between deep characterization and AI is unprecedented. Is there no related work in this space? Or no related work worth mentioning?”

In response, we first note that our firm, explicit focus on the nexus between the craft of character creation, design, and deployment on the one hand, and formal metrics applied to the science of

---

level of FOL, or a finite number of quantifier-rich formulae in SOL. For details, see (Ebbinghaus et al. 1994).

[22] Again, in the case of ACT-R, its scientific program includes what is explicitly set out in (Anderson & Lebiere 2003).

cognitive modeling (via application to cognitive architectures), does indeed seem to be without precedent (for better or worse). That said, there certainly is some very impressive and important work that relates to AI, characters, and art; and this work is in part relevant to our present essay. For example, there is a form of AI invented (as far as we know) by Mateas (2001): **Expressive AI**. As, Mateas puts it, Expressive AI is the marriage of "art practive" with "AI-based practice," in order to build "cultural artifacts." Mateas says that "Expressive AI changes the focus from an AI system as a thing in itself (presumably demonstrating some essential feature of intelligence), to the communication begween author and audience." (Mateas 2001, p. 153) Since ultimately at least one of the main purposes in building three-dimentional characters is to produce narrative that moves audiences in profound ways, Expressive AI is obviously relevant to our purposes herein — yet only to a degree. As we have pointed out, it's mathematically possible for a moving narrative, replete with prose that brings characters to life for readers, to be radomly generated, independent of any cognition whatsoever. Hence it follows that it's mathematically possible that cognitive modeling via a cognitive architecture of an author's mind and notebooks/files is entirely superfluous. But our assumption herein is that the standard practice on the part of human authors is worth examining; and when it's examined (with help from the cognition that Movie Outline and other such tools demand from authors), it becomes clear that, at least with respect to the metric of expressivity and algorithm $\mathcal{R}$, cognitive architectures face a tall order.[23]

8 Conclusion

We conclude that the contemporary first-rate craft of character design, if it measures up to the level of thoroughness called for by sophisticated but AI-less software desgined to facilitate such design, generates expressivity challenges that proponents of cognitive architectures should find quite sobering. Some architectures are clearly not up to these challenges; some perhaps have at least some elements that are somewhat promising (ACT, and presumably descendant ACT-R); and some have a very impressive capacity to handle even natural language (and *a fortiori* background facts) that requires third-order logic (TOL) and quantified epistemic logic. We advise the cognitive modeling community to ascertain and announce which of these three categories their favorite cognitive architecture falls under, and why.

As to future work in our own case, we intend to raise the classification of cognitive architectures, with respect to expressivity (and, subsequently, with respect to equally important measures in the relevant mathematics) to a level of greater rigor, so that such classification is expressed and confirmed by relevant meta-theorems. This future for cognitive modeling is directly in line with the "theorem-anchored" one called for in (Bringsjord 2008).

References

Albus, J. S. (2000), 4-D/RCS Reference Model Architecture for Unmanned Ground Vehicles, *in* 'Proceedings of the 2000 IEEE International Conference on Robotics and Automation', San Francisco, CA.

---

[23] The literary cultural artifacts that Mateas and colleagues have themselves created, by the way, could to a significant degree play the roles we have assigned herein to the characters in *Double-Minded Man*, and those in the drama of Ibsen and fiction of Ullman. However, it's worth noting that there are no formulae corresponding to those at the crux of the expressivity challenge we articulate herein, to be found. E.g., the representations in (Mateas 1997) at the code level are thin, and wouldn't serve to flesh out the requirements in Movie Outline 3 for declarative information about and constitutive of a character (this isn't in the least a criticism).

Anderson, J. (1976), Chapter 7: Formal Semantics of ACT Represetations, *in* 'Language, Memory, and Thought', Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 220–251.

Anderson, J. & Lebiere, C. (2003), 'The Newell Test for a Theory of Cognition', *Behavioral and Brain Sciences* **26**, 587–640.

Anderson, J. R. (1993), *Rules of Mind*, Lawrence Erlbaum, Hillsdale, NJ.

Anderson, J. R. & Lebiere, C. (1998), *The Atomic Components of Thought*, Lawrence Erlbaum, Mahwah, NJ.

Andrews, P. (2002), *An Introduction to Mathematical Logic and Type Theory: To Truth Through Proof*, Springer, New York, NY.

Barwise, J. & Etchemendy, J. (1999), *Language, Proof, and Logic*, Seven Bridges, New York, NY.

Blackburn, P. & Bos, J. (2005), *Representation and Inference for Natural Language: A First Course in Computational Semantics*, CSLI, Stanford, CA.

Bringsjord, S. (1992), CINEWRITE: An Algorithm-Sketch for Writing Novels Cinematically, and Two Mysteries Therein, *in* M. Sharples, ed., 'Computers and Writing: State of the Art', Kluwer, Dordrecht, The Netherlands.

Bringsjord, S. (1995*a*), Could, How Could We Tell If, and Why Should–Androids Have Inner Lives?, *in* K. Ford, C. Glymour & P. Hayes, eds, 'Android Epistemology', MIT Press, Cambridge, MA, pp. 93–122.

Bringsjord, S. (1995*b*), Pourquoi Hendrik Ibsen Est-Il Une Menace pour La Littérature générée Par Ordinateur?, *in* A. Vuillemin, ed., 'Littérature et Informatique la Littérature Générée Par Orinateur', Artois Presses Universite, Arras, France.

Bringsjord, S. (1995*c*), Pourquoi hendrik ibsen est-il une menace pour la littérature générée par ordinateur?, *in* A. Vuillemin, ed., 'Littérature et Informatique la Littérature Générée Par Orinateur', Artois Presses Universite, Arras, France, pp. 135–144.

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, and is available in the online archives of *Game Studies*.
URL: *http://www.gamestudies.org*

Bringsjord, S. (2008), Declarative/Logic-Based Cognitive Modeling, *in* R. Sun, ed., 'The Handbook of Computational Psychology', Cambridge University Press, Cambridge, UK, pp. 127–169.
URL: *http://kryten.mm.rpi.edu/sb_lccm_ab-toc_031607.pdf*

Bringsjord, S. & Bringsjord, A. (2009), Synthetic Worlds and Characters, and the Future of Creative Writing, *in* C. A. P. Smith, K. Kisiel & J. Morrison, eds, 'Working Through Synthetic Worlds', Ashgate, Surrey, England, pp. 235–255.

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

Bringsjord, S., Govindarajulu, N., Ellis, S., McCarty, E. & Licato, J. (2014), 'Nuclear Deterrence and the Logic of Deliberative Mindreading', *Cognitive Systems Research* **28**, 20–43.
URL: *http://kryten.mm.rpi.edu/SB_NSG_SE_EM_JL_nuclear_mindreading_062313.pdf*

Charniak, E. & McDermott, D. (1985), *Introduction to Artificial Intelligence*, Addison-Wesley, Reading, MA.

Cohen, L. H. (2014), 'Suddenly that Summer', *The New York Times Sunday Book Review* p. BR15.

Ebbinghaus, H. D., Flum, J. & Thomas, W. (1984), *Mathematical Logic*, Springer-Verlag, New York, NY.

Ebbinghaus, H. D., Flum, J. & Thomas, W. (1994), *Mathematical Logic (second edition)*, Springer-Verlag, New York, NY.

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

Gall, D. & Frühwirth, T. (2014), A Formal Semantics for the Cognitive Architecture ACT-R, *in* M. Proietti & H. Seki, eds, 'Logic-Based Program Synthesis and Transformation: (Lecture Notes in Computer Science Volume 8981)', Springer, Cham, Switzerland, pp. 74–91.
URL: *http://www.informatik.uni-ulm.de/pm/fileadmin/pm/home/fruehwirth/drafts/act_r_semantics.pdf*

Goble, L., ed. (2001), *The Blackwell Guide to Philosophical Logic*, Blackwell Publishing, Oxford, UK.

Holyoak, K. J. & Hummel, J. E. (2000), The Proper Treatment of Symbols in a Connectionist Architecture, *in* E. Deitrich & A. Markman, eds, 'Cognitive Dynamics: Conceptual Change in Humans and Machines', MIT Press, Cambridge, MA.

Hummel, J. E. & Holyoak, K. J. (1997), 'Distributed Representations of Structure: A Theory of Analogical Access and Mapping', *Psychological Review* **104**(3), 427–466.

Hummel, J. E. & Holyoak, K. J. (2003), 'Relational Reasoning in a Neurally-plausible Cognitive Architecture: An Overview of the LISA Project', *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society* **10**, 58–75.

Jaśkowski, S. (1934), 'On the Rules of Suppositions in Formal Logic', *Studia Logica* **1**, 5–32.

Keisler, H. J. (1986), *Elementary Calculus: An Infinitesimal Approach*, North-Holland, Amsterdam, The Netherlands.

Larson, R. & Edwards, B. (2014), *Calculus (10e)*, Brooks/Cole, Boston, MA. This is an ebook provided by Amazon for their Kindle reader.

Licato, J. (2015), Analogical Constructivism: The Emergence of Reasoning Through Analogy and Action Schemas, PhD thesis, Rensselaer Polytechnic Institute, Troy, NY.

Licato, J., Sun, R. & Bringsjord, S. (2014*a*), Structural Representation and Reasoning in a Hybrid Cognitive Architecture, *in* 'Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN)'.

Licato, J., Sun, R. & Bringsjord, S. (2014*b*), Using a Hybrid Cognitive Architecture to Model Children's Errors in an Analogy Task, *in* 'Proceedings of CogSci 2014'.

Licato, J., Sun, R. & Bringsjord, S. (2014*c*), Using Meta-Cognition for Regulating Explanatory Quality Through a Cognitive Architecture, *in* 'Proceedings of the 2nd International Workshop on Artificial Intelligence and Cognition', Turin, Italy.

Mateas, M. (1997), Computational Subjectivity in Virtual World Avatars, Technical Report FS-97-02, AAAI.
URL: *http://www.aaai.org/Papers/Symposia/Fall/1997/FS-97-02/FS97-02-021.pdf*

Mateas, M. (2001), 'Expressive AI: A Hybrid Art and Science Practice', *Leonardo: Journal of the International Society for Arts, Sciences, and Technology* **34**(2), 147–153.

Newell, A. (1981), 'The Knowledge Level', *Artificial Intelligence* **18**, 18–127.

Partee, B., Meulen, A. & Wall, R. (1990), *Mathematical Methods in Linguistics*, Kluwer, Dordrecht, The Netherlands.

Rosenbloom, P., Newell, A. & Laird, J. (2014), Toward the Knowledge Level in SOAR: The Role of the Architecture in the Use of Knowledge, *in* K. van Lehn, ed., 'Architectures for Intelligence: The 22nd Carnegie Mellon Symposium on Cognition', Psychology Press.

Russell, S. & Norvig, P. (2009), *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ. Third edition.

Sullivan, K. (1974), *The Teaching of Elementary Calculus: An Approach Using Infinitesimals*, Madison, WI. This is a PhD thesis at the University of Wisconsin.

Sun, R. (1995), 'Robust Reasoning: Integrating Rule-Based and Similarity-Based Reasoning', *Artificial Intelligence* **75**(2).

Sun, R. (2001), *Duality of the Mind*, Lawrence Erlbaum Associates, Mahwah, NJ.

Sun, R. (2002), *Duality of the Mind: A Bottom Up Approach Toward Cognition*, Lawrence Erlbaum Associates, Mahwah, NJ.

Sun, R. (2003), A Tutorial on CLARION 5.0, *in* 'Cognitive Science Department Technical Reports', Rensselaer Polytechnic Institute.
URL: *http://www.cogsci.rpi.edu/∼rsun/sun.tutorial.pdf*

Sun, R. & Zhang, X. (2004), Accounting for Similarity-Based Reasoning within a Cognitive Architecture, *in* 'Proceedings of the 26th Annual Conference of the Cognitive Science Society', Lawrence Erlbaum Associates.

Sun, R. & Zhang, X. (2006), 'Accounting for a Variety of Reasoning Data Within a Cognitive Architecture', *Journal of Experimental and Theoretical Artificial Intelligence* **18**(2), 169–191.

Tversky, A. (1977), 'Features of Similarity', *Psychological Review* **84**(4), 327–352.

Ullmann, L. (2014), *The Cold Song*, Other Press, New York, NY. An English version translated from the orginal Norwegian.