11

A Journey along the Borderland: A Critical Approach to Artificial Intelligence-Based Art and Literary Practices

Jichen Zhu and D. Fox Harrell

Digital media open up opportunities for new integrations of art and science. However, the close contact between the multiple cultures in either tradition also unveils fundamental value differences that impose considerable difficulties in performing interdisciplinary work. In this chapter, we identify a new form of this cultural divide in the context of computer art and digital media practices. Next, we identify a growing number of practices that engage the capacity of the computer to abstractly represent data and to process it algorithmically in order to serve expressive, critical, and generalizable purposes. In particular, we explore artificial intelligence-based art and literary practices that actively negotiate the hidden assumptions and push the disciplinary boundaries of both art and science. Finally, we present our AI-based interactive narrative work *Memory, Reverie Machine*, which engages literature, cognitive science, and AI. It illustrates our perspective and strategy of combining art and science practices synergistically, as part of a growing community for which the exploring of the borderland between art and science can transform not only particular technologies or how they are perceived, but also end goals and values.

11.1 Introduction

Half a century ago, British scientist and novelist C. P. Snow [Snow, 1964] delivered a now notorious talk on the increasing gap between the two cultures of the sciences and the humanities [Lam, 2008]. Reflecting on his personal and social experiences as a scientist and novelist, Snow pointed out the lack of communication fueled the deep cultural divide between the two major intellectual campaigns. According to Snow, the resulting misconceptions and distrust was counterproductive to solving real-world problems.

To those who were concerned with Snow's intellectual divide, the advent of digital media brought new opportunities to bridge the gap. Indeed, in the last decades, the humanities/arts and science/engineering communities have developed and shared numerous common approaches to the computer. In areas concerning screen-based content, practitioners constructed metaphors of screens as "pages," widely used in both net art and scientific data archives, or the screen as a "canvas," in animated digital painting and information visualization in computing research. Some see computer as a tool, for example, in the creation of synthesized music as well as the design of a new chip layout. Others may conceive it as something for humans to relate to [Turkle, 2004], in the form of either cajoling video games characters or medical robots that treat ailments and attend to patients' emotional needs. Finally, thinkers and practitioners from both cultures may place their focus on computer algorithms and/or knowledge structures, as the cases in software art and artificial intelligence research.

Despite what may seem like a new Renaissance age, however, the close contact between the arts and sciences[1] does not automatically eliminate the fundamental differences in their values and methodology. For instance, to offer an over-generalized summary, goals of the arts often revolve around issues such as aesthetics, expression, and/or social critique, whereas the sciences still typically value utility, empirical understanding of the world, and/or productively generalizable insights.

We argue that uncritical crossovers between the two cultures may polarize their differences and deepen the divide. In Snow's time, intellectuals ignored each other. "Oh, those are mathematicians! We never talk to them" [Snow, 1964]. If the gap fifty years ago was mainly due to the lack of interest and basic literacy in the other culture, it takes its shape as the battle between conflicting core values and beliefs. Unable to reconcile these differences, many early digital art-combine-with-

[1] We will use the umbrella term "sciences" to refer to the related communities of sciences and engineering, and "arts" for the arts, humanities, and design.

science experiments fell short of their interdisciplinary claim by merely scratching the surface of the other culture. They simply regarded the other cultures as merely a tool for accomplishing work in the domains of their native cultures. This cultural imperialism has intended consequences. For instance, uncritically using simulation technologies without taking into account its deep root in military trainings may undermine ethical claims of an art piece [Penny, 2004]. Similarly, trying to artificially constrain digital artwork to recognizable art world conventions (e.g., hanging framed computer screens as new computer "paintings" and ignoring century-worth of discourse on paintings and recent theory of the unique characteristics of software art as distinct from that tradition) may be a dangerously traditionalist approach.

In this chapter, we examine the mutation of Snow's two cultures in the context of computer-based art and cultural practice and articulate our strategies of combining art and science practices synergistically as peer practices. In particular, we explore artificial intelligence (AI)-based art and literary practices that actively negotiate the hidden assumptions and push the disciplinary boundaries of both art and science. In the rest of the chapter, we will first identify representations of the two cultures in contemporary digital media practice. We will then discuss existing approaches for which the intersections between art and science can transform not only particular technologies or how they are perceived, but also end goals and values. In dialogue with notions such as Michael Mateas's "expressive AI" [Mateas, 2001] and Noah Wardrip-Fruin's "Expressive Processing" [Wardrip-Fruin, 2009], we engage the capacity of the computer to abstractly represent data and to process it algorithmically in order to serve expressive, critical, and generalizable purposes. Next, we will present our AI-based interactive narrative project *Memory, Reverie Machine*, which engages literature, cognitive science, and AI. Finally, we will discuss our strategies of combining art and science practices synergistically.

11.2 Integrating the Two Cultures

Today, the intersection and exchange between the two cultures take place frequently in the realm of digital media. An increasing number of practitioners attempted to reach out to the other side of the divide. However, the distance between the two cultures does not seem to be disappearing as many people predicted. In the rest of this section, we will paint a broad picture of the contemporary manifestations of both campaigns in the context of computer-based art and digital media, and why we found certain attitude towards interdisciplinary work problematic. We will also offer what we consider as more fruitful approaches for engaging and bridging the two sides of the divide.

11.2.1 "Duchamp-Land" and "Turing-Land"

Parallel to Snow's two cultures, digital artist and theorist Lev Manovich [Manovich, 1996] depicted a similar divide in computer arts between what he called *Duchamp-land* and *Turing-land*. The former refers to the community of galleries, major museums, prestigious art journals, whereas the latter describes a more technology-focused art world, exemplified by major venues such as ISEA (Inter-Society for the Electronic Arts), Ars Electronica, SIGGRAPH art shows. In spite of their overlapping interest in computer arts, the two lands pay homage to very different intellectual traditions. Duchamp-land, named after the master of modernist avant-garde art, is a continuation of the traditional fine art practices. It orients itself towards "content" and multiple cultural codes, often with an ironic, self-referential, and destructive attitude[2]. Turing-land, on the other hand, gravitates towards the latest, state-of-the-art technology, and is frequently quite self-reflexive, engaging and exploiting the nature of the technologies themselves.

As computer art evolved, venues became more accommodating in letting in different approaches. But the cultural divide between the arts and sciences persisted. Like Manovich, Simon Penny [Penny, 2007] observed a similar polarity in works that address both computing and the arts across disciplines:

[2] Some destructive art practices that Manovich refers to include the Self-destructive Machines by Tinguely and the first exhibition of Nam June Paik "where he screwed technology --- ripping open television sets or changing TV signals by affixing magnets to the monitors."

> A significant difference between computer science research and media arts practice lies in the ontological status of the artifact… for an artwork, the effectiveness of the immediate sensorial effect of the artifact is the primary criterion for success. It is engaging, it is communicative, it is taken to be coherent, or it is a failure. The criterion for success is performative. Most if not all effort is focused on the persuasiveness of the experience. Backstage may be a mess, a kluge. In computer science the situation is reversed. If the physical presentation is a little rough around the edge, or even missing entire pieces, this can be overlooked with a little handwaving, because the artifact functions as a "proof of concept" which points to the real work, which is inherently abstract and theoretical.

These different approaches and perspectives, Penny argued, are deeply rooted in the core ideologies of the two cultures. Science's insistence upon "alphanumeric abstraction," logical rationality, and desire for generalizability are fundamentally in conflict with the affective power of artwork, which is based on specificity and complexity.

When Snow made his observation, the two cultures would not and could not talk to each other because practitioners well-versed in one rarely had basic level of literacy in the other. Fifty-one years later, computational literacy is much more widespread. Many exciting explorations and collaborations are taking place across the cultural divide, and new interdisciplinary areas are emerging. However, we also need to be aware that the conflict between the two cultural ideologies has grown arguably more intense with this close contact.

Our major concern is that some of these collaborations and interdisciplinary inquiries are motivated by an implicit "cultural imperialism," instead of healthy, informed exchanges. As we pointed out earlier, some practitioners from either side saw the other culture as merely a foreign land that their native culture is set to conquer. Under this mindset, some computer scientists see arts as an application domain to which their algorithmic framework can be applied to; likewise, some artists regard computers as merely a tool to achieve their unaltered visions, without understanding or questioning the worldviews that these "foreign" elements embody. In regard to uncritically adopting ideologies from the traditional science world to computing, and particularly to art practice, Penny warned us of the danger of the "unquestioned axiomatic acceptance of the concept of generality as being a virtue in computational practice … especially when that axiomatic assumption is unquestioningly applied in realms where it may not be relevant." Similarly, attempts to reproduce art as a scientific experiment without an in-depth understanding of its values and discourses are equally simple-minded or even costly.

11.2.2 The Borderland of Critical Computing

Though the phrase has been used previously, Harrell's particular notion of critical computing is in the context of producing what he calls "phantasmal media" — most simply described as those computational works that engage human culture, ideology, and conceptualization through algorithmic and data-structural means [Harrell, 2010]. In short, phantasmal media use computing for subjective, cultural, and critical aims. Regarding critical computing in particular, he states that it:

> … refers to the potential for using algorithmic processing and data structuring as expressive bases for expressing commentary about, and making impactful change upon, the world of human experience. The critical computing concept helps technologists to move beyond development of utilitarian and productivity-oriented applications [Harrell, 2010].

Though here critical awareness is directed both externally toward the world and internally toward technology itself, a major inspiration for this perspective comes from self-critical approaches to computing and information sciences. This is perhaps best exemplified by the work of Philip Agre [1997a; 1997b] that addresses the confrontation between the two intellectual traditions in the contemporary research area of artificial intelligence (AI). As a branch of computer science dedicated to the formal study and production of human-level intelligence through computer algorithmic operations, AI also directly engages many long-standing concerns in the humanities tradition, such as the nature of intelligence and intentionality, and sparked debates in boarder contexts. In the political context of a rising Cold War [Edwards, 1996], its ambitious goal and discursive power not only attracted many computer scientists, engineers, psychologists and military funding agencies, but also quickly engaged philosophers, critical theory scholars, practicing artists, and popular culture producers who were concerned with the nature of intelligence and its implications of human identity. Some multi-disciplinary debates of AI seemed like battles between parallel universes, unable to establish a shared intellectual space and time; but others led to fruitful exchanges and reflections [Suchman, 1987, Weizenbaum, 1976, Winograd & Flores, 1986]. Among them, Agre's *critical technical practice* [Agre, 1997b] has far-reaching impact in other technological fields beyond AI — for example Phoebe Sengers's research applies his ideas to highly original approaches to Human-Computer Interaction (HCI) designs using computing within everyday experiences and environments [Sengers, Kaye, Boehner, Fairbank, Gay, Medynskiy , & Wyche, 2004].

In his attempt to reform the field, Agre turned to the critiques of AI from various humanities fields — phenomenology, literary theory, and anthropology. At first, from the perspective of a more traditional AI practitioner, these texts seemed impenetrable, and hostile. As described by Agre [Agre, 1997b]:

> All critical analysis may seem like accusations of conscious malfeasance. Even sociological descriptions that seem perfectly neutral to their authors can seem like personal insults to their subjects if they presuppose forms of social order that exist below the level of conscious strategy and choice.

After considerable cultural shocks, self-reflection and adaptation, his first breakthrough came when:

> [I]t finally occurred to me to stop translating these strange disciplinary languages into technical schemata, and instead simply to learn them on their own terms. This was very difficult because my technical training had instilled in me two polar-opposite orientations to language — as a precisely formalized and as

> impossibly vague — and a single clear mission for all discursive work—transforming vagueness into precision through formalization… I understood intellectually that the language was "precise" in a wholly different sense from the precision of technical language.

The gradual internalization of these foreign works provided new vocabularies, methods, and perspectives to scrutinize the hidden assumptions and worldviews related to science and engineering, often taken for granted by its practitioners. At the core of his argument, Agre challenged the long-standing premise that science and engineering are objective and value-neutral, and pointed out different "ideologies and social structures [embedded in AI research, that are] … reproduced through a myriad of unconscious mechanisms such as linguistic forms and bodily habits." The goal of a critical technical practice is thus to "cultivate awareness of the assumptions that lie implicit in inherited technical practices" [Agre, 1997a].

Agre's work demonstrates the importance of constructive collaborations between the two cultures. It inspired a new community of critical technical and technical critical practices, including, among others, Sengers's AI agent design informed by culture studies [Sengers, 1998], Michael Mateas's expressive AI [Mateas, 2001], Ian Bogost's procedural rhetoric [Bogost, 2006], Noah Wardrip-Fruin's expressive processing [Wardrip-Fruin, 2009], Harrell's phantasmal media [Harrell, 2009], and Zhu's AI Hermeneutic Network [Zhu, 2009, Zhu & Harrell, 2009].

11.3 Artificial Intelligence, Cognitive Science and Stream of Consciousness

"Stream of consciousness" is a psychological term that William James coined in his 1890 text *The Principles of Psychology* [James, 1890]. The term was later applied to works by various modernist writers such as Dorothy Richardson, James Joyce, Virginia Woolf, and William Faulkner, indicating both their literary techniques and the genre itself. Beyond various formal experiments, stream of consciousness literature reflects a conceptual purpose — to use the internal thoughts as a primary way of depicting fictional characters. As Humphrey puts it in *Stream of Consciousness in the Modern Novel* [Humphrey, 1954], the works under this genre replace the motivation and action of the "external man" with the psychic existence and functioning of the "internal man."

Decades have passed since modernist authors' initial experiments and many works associated with this literary experiment have entered the canon of "high" literature. Their approach for expressing aspects of human subjectivity and pre-speech consciousness nevertheless are still relevant to many recent technologies (e.g., AI), theories (e.g., cognitive linguistics), and forms (e.g., computational narrative). These younger developments, in their own ways, have taken the modernist writers' steps further in ways described below.

In this section, we call attention to underlying parallels and synergies between stream of consciousness literature, cognitive linguistics, and AI, as the motivation of our computational narrative project, described in the next sections. We believe that concerns of modernist writers regarding inner thoughts have been reinvigorated in light of these contemporary cognitive scientific developments regarding preconscious conceptualization. As a critical technical practice, our work in algorithmically narrating characters' memory, reverie, and daydreaming (Section 11.4) exemplifies a new literary form that can leverage AI technologies for expressive narrative without being burdened by the former's philosophical baggage or implicit aesthetic dictates.

11.3.1 *Stream of Consciousness Literature and Artificial Intelligence*

Stream of consciousness writing and AI may pose an unlikely match as a subject of comparative analysis. The two fields not only sprouted in different historical periods, but also reside in two separate communities. One was populated in the early twentieth century and is now associated with academic literary analysis more often than being seen as vibrant area for active creative production, whereas the other is a still on-going development in the techno-science sphere that underwent significant selfreevaluation after the so-called "AI-Winter" of the 1970s [Russell & Norvig, 2002]. Beneath the obvious differences, however, are the similar overarching goals in their respective historical contexts and parallel roles that they both take on in their relationships to contemporaneous concerns.

First, stream of consciousness literature and AI speak to each other through a shared ambition. Humphrey observed that "[t]he attempt to create human consciousness in fiction is a modern attempt to analyze human nature" [Humphrey, 1954]. If stream of consciousness writers sought their answers by portraying humans directly, the AI community pursued theirs by constructing the "other" – machines. AI practitioner Michael Mateas recently echoed that "AI is a way of exploring what it means to be human by building systems" [Mateas, 2002]. These systems, built in attempt to resemble or surpass their human creators, have become our mirrors to reflect upon our identities as humans [Turkle, 1984].

Secondly, both fields rejected behaviorism in their respective historic periods, and turned their attentions to what happens internally in human mental activities as gateways to understanding "human nature." Prior to the turn of the twentieth century, fictional characters were typically represented by their external behaviors. Writers carefully crafted their actions, dialogues, and rational thoughts to create distinctive personas for their stories. What stream of consciousness writers were able to achieve, in comparison, was to create their characters mainly out of their psychological aspects, including their buzzing random thoughts and associative trails.

The scientific community from which AI grew out of in the 1950s, in parallel, was similarly dominated by behaviorism. The paradigm was based on the laws of stimulus-response and declared itself as the only legitimate scientific inquiry. Mental constructs such as knowledge, beliefs, goals and reasoning steps were dismissed as unscientific "folk psychology" [Russell & Norvig, 2002]. Part of AI's contribution was to bring these scientific taboo back to the table by building powerful computational systems based on them. Like Newell and Simon's 1957 *General Problem Solver* [Newell, Shaw & Simon, 1959], many research efforts have been poured into modeling human cognitive capabilities, including reasoning, planning, and learning.

11.3.2 *Stream of Consciousness Literature and Cognitive Linguistics*

The pre-speech level of thought that was neglected by the AI community has been scrutinized again recently in a new field closely built, in part, upon AI: cognitive science. To contemporary cognitive linguists, such as Gilles Fauconnier, George Lakoff, Mark Johnson, and Mark Turner, this neglected land of consciousness holds the basis for our basic conceptual, and even literary thought [Fauconnier, 1985, Fauconnier & Turner, 2002, Lakoff & Johnson, 1980].

Language is only the tip of a spectacular cognitive iceberg, and when we engage in any language activity, be it mundane or artistically creative, we draw unconsciously on vast cognitive resources, call up innumerable models and frames, set up multiple connections, coordinate large arrays of information, and engage in creative mappings, transfers, and elaborations. Beneath the tip of this iceberg is a wide range of cognitive phenomena that Fauconnier calls "backstage cognition," defined as "the intricate mental work of interpretation and inference that takes place outside of consciousness" [Fauconnier, 2001, Fauconnier & Turner, 2002]. Thus, we could say that cognitive linguists cite phenomena that are even below the unarticulated thought phenomena explored by stream of consciousness authors — but at a level that still addresses conceptualization as opposed to perception, motor-action, or other pre-conscious cognitive phenomena.

11.3.3 *Benefits and Challenges*

In the course of working with these very different traditions and methods, we encountered numerous challenges and also compensated by many new perspectives and insights. Some obstacles are a direct result of the clashing differences between the Turing-land and the Duchamp-land. Although both communities were interested in cognitive phenomena, stream of consciousness writers and AI practitioners emphasize different stages of human consciousness. The term "consciousness" from the vantage point of modernist writers referred to "the whole area of mental processes, including especially the pre-speech levels" [Humphrey, 1954].

This was based on James' original psychological theory, in which "memories, thoughts, and feelings exist outside the primary consciousness" and, further, that they appeared, not as a chain, but as a stream, a flow [James, 1890]. Early AI, on the other hand, regarded human rationality as the key to problem solving. Early practitioners in the field relied upon the rational and stable operations of our cognitive processes at the cost of the addressing the roles of the body, affect, and the uncontrollable stream of thoughts unmediated by logic and rationality.

Another conflict is due to the two cultures' opposite values of specificity and generalizability. Modernist writers such as Virginia Woolf believed that the important subject for an artist to express was her private and subjective vision of reality. Woolf's characters all embodied her belief in the individual's constant search of meaning and identification [Humphrey, 1954]. This individualistic approach contrasts strongly with AI's focus on generalizability, in which individual differences are often sacrificed for regularity and scalability. We reconcile these two stances by distancing our work from an attempt to reduce mental activities to uniform formal algorithmic processes. Instead, our project utilizes scientific computational methods, including logical/mathematical formalization, as a way to express our human search for meaning. Formal representations are no more or less meaningful than any other human forms of symbol making; their benefits of precision and computational implementability come at a cost of subjective interpretability.

In the meanwhile, forging the bond across the cultural gap provides unique opportunities. It may be argued that one of the reasons that early AI largely confined itself to the territory of rationality is the extreme difficulty that the field ran into in its attempt to model common sense and contextual reasoning explicitly. These powerful, but for the most part invisible, operations are seen within the field of cognitive linguistics to be partially observable in the structure of our linguistic creations. The insights posed by the cognitive linguistics enterprise and the expressive needs invigorated by our interdisciplinary approach offer the opportunity to revisit some of the compromises that AI made in its early stage.

11.4 Memory, Reverie Machine

The *Memory, Reverie Machine* (MRM) project reflects our approach for invigorating the cultural concerns of depicting internal reflection and imaginative subconscious and for applying cognitive science and AI techniques for expressive purposes. MRM is an interactive narrative system that generates different stories based on user input. The main character of these stories is a robot called Ales who gradually discovers himself. In the beginning of the story, Ales functions more like an avatar and is controlled completely by the user. As the story progresses, however, he starts to recall memories triggered by the artifacts and other characters he interacts with. These memories construct Ales' beliefs and desires and determines his emotional states. In the later stage of the story, if the user commands him to perform actions contradictory to his beliefs and desires, Ales may ignore it. Even if he does it eventually, he will do it very reluctantly. At the end of the story, Ales may gain its full autonomy by acting on its own completely or getting lost in his internal world forever.

The sample output in Fig. 11.1 illustrates one iteration of user interaction with the system, containing content from both the main (objective) story world and Ales' internal memory world. Depending on user input, the system will generate different text with different emotional tones in ways that we will discuss below. For the rest of the section, we describe the influences of forbearers in conjoining concerns of computing and literature, and foundational work by one of the authors for our current project. We then highlight the model proposed by MRM using an example comprised of actual system output.

```
the color of the kitchen door today should be [yellow/blue]
-> blue
ales said to himself, adjusting his visual scan system.
the room where he had his first encounter of tune-up and oil
change
had similar depressing entryways.
the oil change left a sickly feeling in his gut;
he would rust like the tin man before enduring another.
it amazes him that no one else complained about these things;
they were too busy with the kitchen;
he should probably start his daily routine now.
```

Fig. 11.1. Sample output from *Memory, Reverie Machine*. User input is indicated by "->".

11.4.1 *Literary Antecedents*

The goal of stirring human imagination through literary works that are different on each reading is not new. It often involves explorations of creative integrations of mathematical and/or algorithmic concepts and writers' meaningful expressions. An early work is Raymond Queneau's 1961 *Cent Mille Milliards de Poémes* (*One Hundred Thousand Billion Poems*), originally published as a set of ten sonnets with interchangeable lines [Queneau, 1961]. It explores the idea of writing as a combinatorial exploration of possibilities. Another member of the experimental literary group Oulipo, Italo Calvino, in his essay/lecture *Cybernetics and Ghosts* [Calvino, 1982], claimed that writing was a combinatorial game and "the operations of narrative, like those of mathematics, cannot differ all that much from one people to another, but what can be constructed on the basis of these elementary processes can present unlimited combinations, permutations, and transformations." In Calvino's novels such as "If on a winter's night a traveler" there was a strong sense of narrative coherence, yet Calvino also carefully explicated the algorithmic generation of the novel's form [Calvino, 1995].

The introduction of AI to electronic literary works laid a foundation for the natural integration of AI, cognitive science, and literary concerns. One of the first computer story generation systems, Meehan's Tale-Spin [Meehan, 1981] produced simple animal fables, with the goal of exploring the creative potential of viewing narrative generation as a planning problem, in which agents select appropriate actions, solve problems within a simple simulated world, and output logs of their actions. A more recent example is Selmer Bringsjord and David Ferrucci's 2000 BRUTUS system, which aims to explore formalizations for generating stories about betrayal, with the goal of being "interesting" to human readers.

It is worthwhile to call attention to the different approaches in these two computer science-based systems. While Tale-Spin directly exposed a reader to the output of a planning algorithm, BRUTUS used rich textual descriptions to accentuate the narrative outcome, an approach native to the Duchamp-land. In fact, BRUTUS's extensive amount of pre-authored output raised many questions, especially among the computer science community, whether the system actually authored the text. This criticism is only valid if the system is said to be an autonomous author; in our opinion, there is nothing wrong with human authors creating computationally reconfigurable texts

In contrast, William Chamberlain and Thomas Etter's dialogue-based program Ractor, and Ractor's book *The Policeman's Beard is Half Constructed* [Racter, 1984], used syntactic text manipulation to support conversation with users having text input and poetic output. This was not intended as scientific research, but rather as entertainment, with humorous and clever output. As Charles Hartman [Hartman, 1996] stated, it is better not to ask "whether a poet or a computer writes the poem, but what kinds of collaboration might be interesting."

11.4.2 *Conceptual Blending and the GRIOT System*

The cognitive semantics theory of conceptual blending and the GRIOT system are the foundations of the MRM project. In contrast to the notion of computational generativity, the human capacity to generate concepts and metaphors has been explored by cognitive scientists as the root of our literary mental processes. Conceptual blending theory, building upon Gilles Fauconnier's mental spaces theory [Fauconnier, 1985] and elaborating insights from metaphor theory [Lakoff & Johnson, 1980, Turner, 1996], describes the means by which concepts are integrated. Simple examples of blending in natural language are reflected in the mental processes triggered by words like “houseboat” and “roadkill,” and phrases like “artificial life” and “computer virus.” In short, the theory describes how we arrive at new concepts through blending partial and temporary pieces of information. Most importantly, the theory proposes that conceptual blending processes occur uniformly in pre-conscious in everyday thought and in more complex abstract thought such as in literary arts or rhetoric [Fauconnier & Turner, 2002].

The GRIOT system, a platform for implementing phantasmal media in the form of generative and interactive multimedia works, is the foundation of MRM both in terms of technical implementation and our approach to computational narrative [Harrell, 2006, 2007a]. This subsection, adapted from the abstract of [Harrell, 2007b], serves as a high level overview of this perspective, which emphasizes computational narrative works with the following characteristics: generative content, semantics-based interaction, reconfigurable narrative structure, and strong cognitive and socio-cultural grounding. A system that can dynamically compose media elements (such as procedural computer graphics, digital video, or text) to result in new media elements can be said to generate content.

GRIOT’s generativity is enabled by blending-based concept generation as described above. It uses Joseph Goguen’s theory of algebraic semiotic approach from computer science to formalize key aspects of conceptual blending theory [Goguen, 1998]. Technical details can be found in [Goguen & Harrell, 2004]. Semantics-based interaction here means that

1. media elements are structured according to the formalized meaning of their content, and
2. user interaction can affect content of a computational narrative in a way that produces new output that is “meaningfully” constrained by the system’s author.

More specifically, "meaning" in GRIOT indicates that the author has provided formal descriptions of domains and concepts to either annotate and select or generate media elements and subjective authorial intent.

Meaning can also be reconfigured at the level of narrative discourse. The formal structure of a computational narrative can be dynamically restructured, either according to user interaction, or upon execution of the system as in the case of narrative generation. Discourse structuring is accomplished using an automaton that allows an author to create grammars for narratives with repeating and nested discourse elements, and that accept and process user input. Appropriate discourse structuring helps to maintain causal coherence between generated blends. Strong cognitive and socio-cultural grounding here implies that meaning is considered to be contextual, dynamic, and embodied. The formalizations used derive from, and respect, cognitive linguistics theories with such notions of meaning. Using semantically based approach, a cultural producer (someone producing expressive works, though potentially not self-defined as an "artist") can implement a range of culturally specific or experimental narrative structures. In the subsection below we describe our new work that arises from the historical context presented above and the theoretical and creative framework just described.

The goal with GRIOT is quite different from passing a type of Turing test for autonomous creative competence, as described earlier this section. It is designed to provide a technical framework for humans to provide rich content; narrative systems created with GRIOT are meant as cultural products themselves (as opposed to instances of output of such poetic systems). The GRIOT system utilizes models for cognitive science, informed by the cognitive linguistics enterprise's skepticism of regarding the possibility of a formal account of human thought and language such as in [Lakoff & Johnson, 1999], toward expressive ends that are often literary.

11.4.3 *Framework for Memory, Reverie Machine*

So far, we have situated our work in a historical context where stream of consciousness literature, artificial intelligence discourse, and cognitive science research complement each other, as well as the technical framework of this project. In this subsection, we present MRM, a text-based computational narrative project through one of our early results (illustrated in Fig. 11.2 below). For this work it is important to distinguish between the project's various levels of technological and expressive investigation:

1. the system as an abstract model for how computational narratives can be made generative, extensible, and reconfigurable,
2. the system that generates the story,
3. the narration techniques developed,
4. the story content, and
5. each instance of output.

The emphases in this chapter are upon levels 1 and 2, which comprise our technological framework and secondarily on level 3, the narrative techniques to depict inner thoughts, and level 4 the self-reflexive subject matter. Level 3 is influenced by Virginia Woolf's stream of consciousness novel *Mrs. Dalloway* [Woolf, 2002 (1925)]. Below we highlight particular aspects of the system relevant to the algorithmic narration of inner-thought; complete technical information on GRIOT and of MRM can be found in [Zhu & Harrell, 2008].

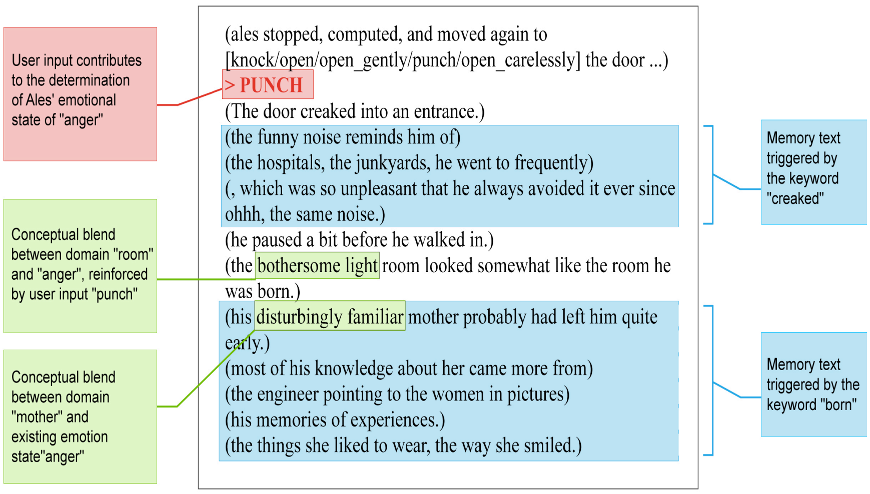

Fig. 11.2. An example of output from Memory, Reverie Machine.

1. Dynamic narration of affect using the Alloy conceptual blending algorithm

Computationally, our system draws upon the GRIOT framework (Section 11.4.2), whose primary generative component is the Alloy algorithm. When modestly applied in MRM, Alloy generates blends involving connecting the main character's current experiences of events, objects, and other characters with affective concepts determined by his current emotional state (see next subsection). In the example in Fig. 11.2, logical axioms selected from an ontology (semantically structured database) describing the concept "room" are blended with axioms describing affective concept "anger," resulting a "bothersome light" room.

The Alloy algorithm uses a set of formal optimality criteria to determine the most common-sense manner in which the concepts should be integrated [Harrell, 2007b]. The result is a blended axiom or set of axioms that is then mapped to natural language output. For example, the description of the "door" in the "anger" state may range from "distasteful wood-colored" to "irritatingly sturdy" or more depending on the concepts being blended. Since blending refers to the conceptual integration of multiple concepts, it is important to be clear that blending is not the mere concatenation of words to form compound phrases. In this case, compound phrases, some of the simplest indicators of conceptual blends, are the final result of an underlying process that is semantic, not lexical.

Constructing blends between objective and affective concepts allows us to achieve a balance between author-determined plot and variable theme or emotional tenor. An artifact required by the plot can be depicted in various ways based on the character Ales' internal emotional state. The highly subjective description, in turn, portrays personality traits of the character, a recurrent technique in *Mrs. Dalloway*.

2. The emotion state machine

Actions taken by a character in a computational narrative, which are usually (but not exclusively) selected by a user, can guide building up of a profile that describes user's preferences, history of actions, and analysis of trends in those actions. A quite simple, but effective, way to do this is tracking tokens representing the emotional state of a character based upon actions that the character has taken. MRM allows user to directly influence the emotion state of Ales, and hence the selection of affective concepts for blends. She may choose among an array of pre-defined actions, such as seeing objects as "red," "yellow," "blue" or another color in the fictional robot character's optic sensors, each connecting to a particular emotion. A keyword "red," for instance, may trigger an affective concept "anger." These emotional mappings are designed aesthetically by the authors to achieve narrative effects, not as an attempt at cognitively modeling emotion using computers as in multiple traditional AI projects.

A successful interactive narrative, however, requires a careful balance between the user's agency and author's intention. In our system, user's impact on the character's emotion is moderated by the emotion state machine component for the sake of narrative consistency. The state machine records Ales' current emotion based on the entire history of user input, instead of the most immediate one. It guarantees that changes of Ales' emotions will be gradual, even if user input oscillates between opposite emotions.

3. Memory structuring and retrieval

The GRIOT system is not limited to producing narrative discourse; indeed it has been used for various forms of poetry [Goguen & Harrell, 2004, Harrell, 2007a], for the interactive, generative composition of animated imagery [Chow & Harrell, 2008] and digital video, photography, and illustration[3]. In the case of MRM, we seek to make output coherently extensible at runtime. For this project we allow the narrative to be punctuated with episodic remembered events and longer

---

[3] One of these artworks is titled "Authoring the Living Liberia Fabric: A Generative and Interactive Narrative for Peace, Truth, and Reconciliation," which is shown in the juried exhibition at the 2010 Electronic Literature Organization Conference in Providence, RI. (Author: D. Fox Harrell, Michael Best, Hank Blumenthal, Ayoka Chenzira, Christopher Gonzalez, Andrew Roberts, Natasha Powell, Deji Fajebe, Jason Lee, Paul O'Neil, and Arjun Tomar.)

reveries of remembered experience. Again, this is not meant as an experiment in cognitive modeling or advanced algorithmic design, our goal is to demonstrate discourse that is meaningfully reconfigurable to serve an author's expressive goals in dialogue with a user's selected actions. In MRM, each memory is annotated based on its subject matter and is retrieved when at least one subject item appears in the story line. In the example in Fig. 11.2, Ales' unpleasant memory of hospitals and junkyards is triggered by the opening of a door through the mutual subject of a certain sound. The system also keeps track of the emotional tone of each memory and selects a memory only if it does not clash with the current emotion state. The example in Fig. 11.1 illustrates this feature.

MRM represents an approach that is different from Duchamp-land or Turing-land, explained in Section 11.2, for neither culture is seen as subordinate to the other. We see our approach closely related to "bootstrapping," a terminology in computing that describes how simple programs can build up larger ones. In our project, the exchange between the two cultures strengthens both: new insights in cognitive science theory and discoveries in AI techniques help us to further steer and articulate our expressive goal; renewed aesthetic needs guide the next iteration of algorithmic exploration. In our case, this tightly coupled feedback loop led to tremendous improvements in both aspects.

11.5 Conclusion

Digital technology has no doubt created new borderlands where the arts and sciences intersect: English scholars apply digital statistical tools to hundreds of books and identify patterns and trends beyond the human scale [4]; dancers use computer vision technology to mediate their performances [Nahrstedt, Bajcsy, Wymore, Sheppart , & Mezur, 2008]; and computer scientists use social theory to inform their multi-agent AI systems [McCoy & Mateas, 2009], and more. Yet, despite the growing

[4] This approach, for example, is illustrated by the various works in the "Computational approaches to textual variation in medieval literature" Panel in the 2010 Digital Humanities Conference, held in London, UK.

interest and knowledge between the two cultures, Snow's cultural divide still persists.

We believe these interdisciplinary collaborations can be further extended if we can take advantage of each discipline's strength. More importantly, one can use the other culture as a mirror to make visible and reflect upon the hidden value system and ideologies in his/her practice. In the case of our own interactive narrative project Memory, Reverie Machine, we intend to forge a unique bond: engaging stream of consciousness literature and critically examining subjective experiences of AI technologies without taking a totalizing, modernist stance toward literary production, and engaging cognitive science and AI as a critical technical practice where both the methods used and interactive output are seen as expressive.

While confronting the other culture, Agre [Agre, 1997b] noted the importance of "[m]aintaining constructive engagement with researchers whose substantive commitments I found wildly mistaken. It is tempting to start explaining the problems with these commitments in an alien disciplinary voice, invoking phenomenology or dialectics as an exogenous authority, but it is essentially destructive." It is with this constructive approach — looking not only at the "wildly mistaken", but also the wildly inspiring — that we, along with our peer practitioners in diverse fields, explore the borderland between art and science. We hope that others might engage our work as at least "wildly mistaken" in interesting, useful, and productive ways. More hopefully, it can be seen as a productively "wild" reconciliation of disciplinary values and expressive practice as we forge ahead in our hybrid research/art.

References

Agre, P. E. [1997a] *Computation and Human Experience* (Cambridge University Press, Cambridge, UK).

Agre, P. E. [1997b] "Toward a critical technical practice: Lessons learned in trying to reform AI," in *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide,* eds. Bowker, G. C., Star, S. L., Turner, W. & Gasser, L. (Lawrence Erlbaum Associates, Mahwah, NJ) pp. 131-158.

Bogost, I. [2006] *Unit Operations: An Approach to Videogame Criticism* (MIT Press, Cambridge, MA).

Calvino, I. [1982] "Cybernetics and ghosts," in *The Uses of Literature* (Harvest, Orlando, FL) pp. 3-27.

Calvino, I. [1995] "How I wrote one of my books," in *Oulipo Laboratory: Texts from the Bibliotheque Oulipienne,* eds. Queneau, R. (Atlas, London).

Chow, K. K. N., & Harrell, D. F. [2008] "Generative visual renku: Linked poetry generation with the GRIOT system". *Hyperrhiz: New Media Cultures*,**6**, http://www.hyperrhiz.net/hyperrhiz06/19-essays/79-generative-visual-renku.

Edwards, P. N. [1996] *The Closed World: Computers and the Politics of Discourse in Cold War America* (MIT Press, Cambridge).

Fauconnier, G. [1985] *Mental Spaces: Aspects of Meaning Construction in Natural Language* (MIT Press/Bradford Books, Cambridge, MA).

Fauconnier, G. [2001] "Conceptual blending and analogy," in *The Analogical Mind: Perspectives from Cognitive Science,* eds. Gentner, D., Holyoak, K. J. & Kokino, B. N. (MIT Press, Cambridge, MA).

Fauconnier, G., & Turner, M. [2002] *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities* (Basic Books, New York).

Goguen, J. [1998] "An introduction to algebraic semiotics, with applications to user interface design," in *Proceedings of Computation for Metaphors, Analogy, and Agents*, 1998, Yakamatsu, Japan.

Goguen, J., & Harrell, D. F. [2004] "Style as a choice of blending principles," in *Proceedings of the 2004 AAAI Fall Symposium in Style and Meaning in Language, Art, Music and Design*, 2004, Washington D.C.

Harrell, D. F. [2006] "Walking blues changes undersea: Imaginative narrative in interactive poetry generation with the GRIOT system," in *Proceedings of AAAI 2006 Workshop in Computational Aesthetics: Artificial Intelligence Approaches to Happiness and Beauty*, 2006, Boston, MA.

Harrell, D. F. [2007a] "GRIOT's tales of haints and seraphs: A computational narrative generation system," in *Second Person: Role-Playing and Story in Games and Playable Media,* eds. Wardrip-Fruin, N. & Harrigan, P. (MIT Press, Cambridge, MA).

Harrell, D. F. [2007b] *Theory and Technology for Computational Narrative: An Approach to Generative and Interactive Narrative with Bases in Algebraic Semiotics and Cognitive Linguistics* (Ph.D. Dissertation, University of California, San Diego, La Jolla).

Harrell, D. F. [2009] "Toward a theory of phantasmal media: An imaginative cognition- and computation-based approach to digital media". *CTHEORY*, September 2, 2009, http://www.ctheory.net/articles.aspx?id=610.

Harrell, D. F. [2010] "Toward a theory of critical computing: The case of social identity representation in digital media applications". *CTHEORY*, May 13, 2010, http://www.ctheory.net/articles.aspx?id=641.

Hartman, C. O. [1996] *Virtual Muse: Experiments in Computer Poetry* (Wesleyan, Hanover, CT).

Humphrey, R. [1954] *Stream of Consciousness in the Modern Novel* (University of California Press, Berkeley, CA).
James, W. [1890] *The Principles of Psychology* (Henry Holt, New York).
Lakoff, G., & Johnson, M. [1980] *Metaphors We Live By* (University of Chicago Press, Chicago).
Lakoff, G., & Johnson, M. [1999] *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought* (MIT Press, Cambridge, MA).
Lam, L. [2008] "Science matter: A unified perspective," in *Science Matters: Humanities as Complex Systems,* eds. Burguete, M. & Lam, L. (World Scientific, Singapore) pp. 1-38.
Manovich, L. [1996]. "The death of computer art." (*Rhizome*. http://www.manovich.net/text/death.html, Oct 6, 2010)
Mateas, M. [2001] "Expressive AI: A hybrid art and science practice," *Leonardo*, **34**(2), 147-153.
Mateas, M. [2002] *Interactive Drama, Art, and Artificial Intelligence*. (Ph.D. dissertation, Carnegie Mellon University, Pittsburgh PA).
McCoy, J., & Mateas, M. [2009] "The computation of self in everyday life: A dramaturgical approach for socially competent agents," in *Proceedings of AAAI 2009 Spring Symposium on Intelligent Narrative Technologies II*, 2009, Stanford, CA.
Meehan, J. [1981] "TALE-SPIN," in *Inside Computer Understanding: Five Programs Plus Miniatures,* eds. Riesbeck, C. K. (Lawrence Erlbaum Associates, New Haven, CT).
Nahrstedt, K., Bajcsy, R., Wymore, L., Sheppart, R., & Mezur, K. [2008] "Computational model of human creativity in dance choreography," in *Proceedings of AAAI 2008 Spring Symposium on Creative Intelligent Systems*, 2008, Stanford, CA.
Newell, A., Shaw, J. C., & Simon, H. A. [1959] "Report on a general problem-solving program," in *Proceedings of the International Conference on Information Processing*, 1959, pp. 256-264.
Penny, S. [2004] "Representation, enaction, and the ethics of simulation," in *First Person,* eds. Wardrip-Fruin, N. & Harrigan, P. (MIT Press, Cambridge, MA).
Penny, S. [2007] "Experience and abstraction: The arts and the logic of machines," in *Proceedings of perthDAC 2007: 7th Digital Arts and Culture Conference*, 2007, Perth, Australia.
Queneau, R. [1961] *Cent mille milliards de poémes* (Gallimard, Paris, France).
Racter. [1984] *The Policeman's Beard Is Half Constructed* (Warner Books, New York).
Russell, S., & Norvig, P. [2002] *Artificial Intelligence: A Modern Approach* (2nd ed.) (Prentice Hall, Upper Saddle River, NJ).
Sengers, P. [1998] *Anti-Boxology: Agent Design in Cultural Context*. (Ph.D. Dissertation, Carnegie Mellon Universit, Pittsburgh).

Sengers, P., et al. [2004] "Culturally embedded computing," *Pervasive Computing*, **3**(1), 14-21.

Snow, C. P. [1964] *The Two Cultures* (Menton Books., New York).

Suchman, L. [1987] *Plans and Situated Actions: The Problem of Human-Machine Communication* (Cambridge University Press, Cambridge, UK).

Turkle, S. [1984] *The Second Self: Computers and the Human Spirit* (Simon and Schuster, New York).

Turkle, S. [2004] "Whither psychoanalysis in computer culture?" *Psychoanalytic Psychology*, **21**, 16-30.

Turner, M. [1996] *The Literary Mind: The Origins of Thought and Language* (Oxford University Press, Oxford).

Wardrip-Fruin, N. [2009] *Expressive Processing: Digital Fictions, Computer Games, and Software Studies* (MIT Press, Cambridge, MA).

Weizenbaum, J. [1976] *Computer Power and Human Reason* (Freeman, New York).

Winograd, T., & Flores, F. [1986] *Understanding Computers and Cognition* (Ablex, New York).

Woolf, V. [2002 (1925)] *Mrs. Dalloway* (Harcourt, New York).

Zhu, J. [2009] *Intentional Systems and the Artificial Intelligence Hermeneutic Network: Agency and Intentionality in Expressive Computational Systems*. (Ph.D. Dissertation, Georgia Institute of Technology, Atlanta, GA).

Zhu, J., & Harrell, D. F. [2008] "Daydreaming with intention: Scalable blending-based imagining and agency in generative interactive narrative," in *Proceedings of AAAI 2008 Spring Symposium on Creative Intelligent Systems*, 2008, Stanford, CA, pp. 156-162.

Zhu, J., & Harrell, D. F. [2009] "The artificial intelligence (AI) hermeneutic network: A new approach to analysis and design of intentional systems," in *Proceedings of the 2009 Digital Humanities Conference*, 2009, pp. 301-304.

D. Fox Harrell is a researcher exploring the relationship between imaginative cognition and computation. He is an Assistant Professor of Digital Media at the Georgia Institute of Technology. He directs the Imagination, Computation, and Expression [ICE] Lab/Studio (icelab.lcc.gatech.edu) in developing new forms of computational narrative, gaming, social networking, and related technical-cultural media based in computer science, cognitive science, and digital media arts. The National Science Foundation has recognized Harrell with an NSF CAREER Award for his project "Computing for Advanced Identity Representation." Harrell holds a Ph.D. in Computer Science and Cognitive Science from the University of California, San Diego. He also earned a B.F.A. in Art, a B.S. in Logic and Computation, and minor in Computer Science at Carnegie Mellon University, each with highest honors. He has also worked as an interactive television producer and as a game designer. *Email: fox.harrell@lcc.gatech.edu*.

Jichen Zhu is an assistant professor of Digital Media in School of Visual Arts and Design at University of Central Florida, where she is the director of the Procedural Expression Lab. Her work focuses on developing humanistic and interpretive theoretical framework of computational technology, particularly artificial intelligence (AI), and constructing AI-based cultural artifacts. Her current research areas include digital humanities, software studies, computational narrative, and serious games. Jichen Zhu received a Ph.D. in Digital Media and a M.S. in Computer Science from Georgia Institution of Technology. She also holds a Master of Entertainment Technology from Carnegie Mellon University and a B.S. in Architecture from McGill University in Montreal, Canada. *Email: jzh@mail.ucf.edu.*