The Pleasures of Immersion and Engagement:
Schemas, Scripts, and the Fifth Business

Abstract
Presently, designers of interactive narratives and video games have only a slender understanding of the aesthetic experiences their audiences and users seek. Using schema theory, this study articulates the two varieties of aesthetic pleasures that users of interactive works enjoy: immersion and engagement. It uses schema theory to define the characteristics of immersion and engagement in both conventional and new media. After examining how readers' experiences of these two different aesthetics may be enhanced or diminished by interface design, options for navigation, and other features, the essay concludes by looking beyond immersion and engagement to "flow," a state in which readers are both immersed and engaged.

It must be granted that there is some value in mystification, labyrinth, or surprise in the environment... This is so, however, only under two conditions. First, there must be no danger of losing basic form or orientation, of never coming out. The surprise must occur in an overall framework; the confusions must be small regions in a visible whole. Furthermore, the labyrinth or mystery must in itself have some form that can be explored and in time apprehended. Complete chaos without hint of connection is never pleasurable. *Kevin Lynch, The Image of the City*

Those roles which, being neither those of Hero nor Heroine, Confidante nor Villain, but which were nonetheless essential to bring about the recognition or the dénouement, were called the Fifth Business in drama and opera companies organized according to the old style; the player who acted these parts was often referred to as Fifth Business. *Robinson Daines, The Fifth Business: The Deposed Trilogy*

Introduction: Schemas and the Pleasure Principle
Oddly enough, after decades of theorizing about texts, their authors (or lack thereof), and their relationship to readers, economics, and culture, we know comparatively little about the affective pleasures of reading. Why do we read for pleasure? What keeps us turning the pages between an author's name on the title page and a novel's last gasp on its very last page? How are we able to turn a sprinkling of abstract symbols on a white page into sce­

Schemas enable us to perceive objects and occurrences around us and to make efficient sense of them by consulting our ready-made store of similar occurrences and understandings, which we gain from reading, personal experience, and even advice we receive from others. Schemas may be as simple as the series of understandings and actions that enable us to both recognize what a car is and how to drive one, or as complex as our understanding of the specific roles characters play in, say, teen slasher flicks, where we expect basements to be hotbeds of horror or peril and teenagers having sex to become victims. Schemas also entail scripts, sets of tasks, or actions appropriate to certain
schemas. In Schank’s famous “restaurant script,” people entering restaurants understand, seemingly automatically, what to do with a menu, how to order, and how to behave throughout the restaurant experience.” Scripts, moreover, are flexible, as we can rapidly and easily modify existing scripts to accommodate new scenarios. A single restaurant script easily covers a visit to McDonalds, Le Cirque, a sushi bar, even a Roman antipasti restaurant, where you merely help yourself to whatever dish is being handed around.

Once we have identified a single schema, we begin calling on relevant scripts that shape our perception, navigation, and interaction within a scenario, whether it exists in life, on a page, or in a stroll on the decks of Titanic in the interactive game “Titanic: Adventure Out of Time.” We watch the trials and travails of a couple in a romantic comedy with considerably less trepidation than we would the eponymous couple in the likes of “Romeo and Juliet,” which we know to be a tragedy, because our schemas for romantic comedies tell us that, in the end, the obstacles exist merely to make the final union all the more satisfying. We know we must treat all clues as potentially relevant when we read a Patricia Cornwell mystery, just as our schemas for mysteries also tell us that the author will dangle the central puzzle until the book’s ultimate pages.

Out of Time.”

Scripts, moreover, are flexible, as we can rapidly and easily modify existing scripts to accommodate new scenarios. A single menu, how to order, and how to behave throughout the restaurant help yourself to whatever dish is being handed around.

“Jacob’s Ladder” ends with an “oh-it-was-all-a-dream” schema that accounts for the illogical and unexpected narrative developments in this movie about an extra-textual perspective on the text itself, as well as on the schemas that have shaped it and the scripts operating within it.

Once we have identified a single schema, we begin calling on relevant scripts that shape our perception, navigation, and interaction within a scenario, whether it exists in life, on a page, or in a stroll on the decks of Titanic in the interactive game “Titanic: Adventure Out of Time.” We watch the trials and travails of a couple in a romantic comedy with considerably less trepidation than we would the eponymous couple in the likes of “Romeo and Juliet,” which we know to be a tragedy, because our schemas for romantic comedies tell us that, in the end, the obstacles exist merely to make the final union all the more satisfying. We know we must treat all clues as potentially relevant when we read a Patricia Cornwell mystery, just as our schemas for mysteries also tell us that the author will dangle the central puzzle until the book’s ultimate pages.

“Jacob’s Ladder” ends with an “oh-it-was-all-a-dream” schema that accounts for the illogical and fantastic nature of events by situating them inside a dream or, in this case, a dying man’s last hallucinations.

**Positive: Immersion and Engagement**

When aesthetic objects invite us to rely on certain schemas, they are not, however, necessarily guaranteeing us an entirely predictable experience. Schemas may enhance our pleasure in, say, reading a John Grisham paperback, because they provide a detailed framework that frees us to focus intently on the minutiae of the narrative by providing us information about what, roughly, to expect from the characters, events, and plot generally, as well as, of course, its eventual outcome. ‘The presence and nature of schemas in a work, moreover, dictate not only the type of genre the work belongs to, but also both the sort of audience the work attracts and the kind of affective experience that audience may expect. Genre fiction generally hews tightly to highly normative schemas, while postmodern novels tend to invert narrative conventions and rupture the stock developments and resolutions of mainstream fiction. Not surprisingly, the predictability afforded by genre schemas makes them ideal fodder for the trance-like reading cognitive psychologists like Victor Nell, note is the hallmark of the immersive reading experience. And, of course, as we might expect, immersive affective experiences also tend to garner the largest audiences, as readers pursue immersion to temporarily escape the stresses of everyday life or vicariously enjoy the exploits of fictional characters as an antidote to the mundanity of their own lives. Contrary to expectations, however, our immersion in what some critics might scoff at as “light reading” stems from the steady, unbroken rhythm of our reading, which fully occupies our cognitive capacities. Conversely, readers plowing through more demanding texts, works by what Robert Coover has dubbed “difficult” writers, enjoy no such spell, as the cognitive demands of grappling with the text tend to be discontinuous, involving shifting between competing schemas, prospecting and retrospecting through the text, and pausing over obscure passages. Highly normative schemas enable readers to “lose” themselves in the text in what we might call an immersive affective experience. When immersed in a text, reader’s perceptions, reactions, and interactions all take place within the text’s frame, which itself usually suggests a single schema and a few definite scripts for highly directed interaction. Conversely, in what we might term the engaged affective experience, contradictory schemas or elements that defy conventional schemas tend to disrupt readers’ immersion in the text, obliging them to assume an extra-textual perspective on the text itself, as well as on the schemas that have shaped it and the scripts operating within it.

**Immersive Interactives:**

**Beyond Shoot Outs and Hunt-Quests**

[T]oday’s most successful interactive artists ultimately see interactivity as an evolutionary (rather than revolutionary) step for storytelling. Brent Hurtig, “The Plot Thickens”

Not surprisingly, the earliest digital interactives—video games—drew their cues heavily from a singular schema, turning the early commercial computer games into jazzed-up versions of video arcade games. Whether by accident or design, early game developers hit digital pay dirt by founding their first ventures on the bedrock of two essentials: a recipe for interaction that all but guaranteed a deeply immersive experience and strong, normative schemas borrowed from already-familiar forms of entertainment. The history of invention is, after all, littered with dazzling innovations that either withered rapidly into obscurity or, at best, hibernated for decades before their eventual adoption, mostly due to the object’s very newness. Inventions that are discontinuous with earlier devices and tools tend to offer users few familiar schemas. The fax, phonograph, answering machine, tape player,
and VCR all languished for decades before becoming household fixtures, largely because each of these inventions required users to develop new schemas to accommodate them. Conversely, technologies like Edison’s incandescent light were adopted at the technological equivalent of light-speed—despite potentially crippling problems with the limitations of wiring and distribution of centralized electricity—almost entirely because the new innovations essentially invited users to rely on long-familiar, comfortable schemas and scripts derived from the very forms of technology the innovations were designed to replace. Ironically, innovations seem to be adopted most rapidly when their newness is domesticated, so to speak, by design features that invite users to treat the new object as if it were merely an extension—albeit an improved one—of a familiar object or device. Early video games like Pong stuck to the simple, rigid schema of a ball game with the ball batted between players or against walls. Later successful video games drew off arcade staples that involved escaping through mazes—an approach drawn loosely from the pinball schema—or raining bullets on would-be protagonists, a schema drawn from that staple of county fairs everywhere, the shooting gallery. The result: a game that imposed rigid rules, drawn from already familiar games which could thus be immediately grasped by users, a game featuring fresh local details like gobbling mouths or souped-up weapons, requiring a steady rhythm of interaction. Ironically, the reader paging through Balzac or Dickens, or, for that matter, Judith Krantz, has entered into roughly the same immersive state, enjoying the same high, continuous cognitive load as the runty kid firing fixedly away at Space Invaders.

Later generations of video games have colonized the same turf with notable success. Both Sonic the Hedgehog and Super Mario Brothers, for example, drew off familiar arcade schemas. Yet video and PC game designers have encountered difficulty whenever they have attempted to stray into territory where no dominant schemas reign. Chief among any designers difficulties is how to invite users to interact with the text itself. While the shoot-out is always immediately comprehensible, that particular schema doesn’t easily offer designers sufficient local details to completely differentiate their latest game from the droves of other shoot-outs that have preceded it. Perhaps not terribly surprisingly, game designers have mostly mined only a single other schema: the treasure-hunt-cum-grail-quest, familiar to users of “Myst,” “Titanic,” “Grim Fandango,” and, for that matter, any other interactive that doesn’t require users to shoot anything that moves. Yet the hunt-quest remains a remarkably hardy genre, as its schema, unlike the shoot-out, permits a wealth of local detail, sufficiently rich that its users can become immersed in grappling with both its intricacies and what to do with it.

**Scripts, Voice-Overs, and the Fifth Business**

In immersive interactivity like “Myst” or Shannon Gilligan’s “Virtual Murder” series, our pleasure stems from our ability to discern a single schema and the several scripts it offers us for both interpretation and directed action. Before we so much as glimpse the title screens of your typical hunt-quest, we already know we need to listen intently, collect everything we can lay our mitts on, and put together our tools and clues to solve the local challenges that confront us, which, in turn, will enable us to solve the interactive’s grand challenge—usually something on the order of liberating a prisoner, altering the course of history, or saving the planet. Still, designers have no such clear choice of scripts for interaction. We all understand that guns or knives are essential to shoot-outs, but no such clear scripts exist for actions during the hunt-quest. Furthermore, how do you indicate readiness for action or differentiate the items your protagonist must collect from the normal detritus that makes an environment look convincingly realistic? Contrary to some theorist’s belief, the existence of tools intra-frame or extra-frame does little to disrupt the user’s immersion in the interactive. For example, solving a puzzle, pursuing a clue, or surviving a knife fight involves action that can potentially spill outside the narrative’s frame, as in the multiple-choice replies “Titanic” offers as responses to characters conversations with you or the inventory of items protagonist Robert Cath possesses in “The Last Express.” But the aesthetic remains largely immersive as long as the story, setting, and interface adhere to a single schema.

Users face further interface challenges from narratives like “Last Express” that attempt to stray from reliable gaming conventions that govern the user’s actions, mostly decoding puzzles and dismembering enemies, all of which may frustrate more than engage users expecting well-defined scripts and a tight framework for directed action. In “Titanic,” for example, if you fail to keep an assignation or meander through the ship, the narrative’s clock-time halts abruptly, and all characters vanish, save your steward, who appears periodically to throw you out of the First Class Smoking Lounge or Verandah Café. No agents, frames, or tools exist to jump-start the narrative again once you neglect to pick up the right clues. Even with its highly normative hunt-quest schema, “Myst” frustrates users searching vainly for clues into acts of desperate, random thrashing with the cursor on shrubbery, sundials, library walls, anything that looks like a candidate for the next puzzle challenge. But even with better indicators for interaction, like “The Longest Journey’s” palette of annotated cursor and potential actions (eye, mouth, hand) that flag your potential actions with solid or broken lines and a variety of colors, users are simply left with more elegant weapons to thrash the simulated environment with. If the environment seems particularly bereft of clues or you don’t happen to hit on the ingenious and incredibly improbable notion of using a twig and dinosaur scale to create a funnel, as in “Longest Journey,” immersion evaporates, and you’re left trying to vainly intuit what on earth the designers had in mind when they created the particular scene you’re presently trapped in.
If anything, video game designers, more than their PC counterparts, have even thornier dilemmas awaiting them in interface design during their forays into territory outside the shoot and hunt-quest genres. With “Sydney 2000,” Dreamcast users can train athletes, expose them to some extra coaching, test them in qualifying trials, and, finally direct them to compete in the Olympics. The problem for users and designers alike: how to provide the means for a potentially complex series of interactions using the same controller originally created for shoot-outs. The solution: users must toggle maniacally between the former “shoot” switches to provide athletes with the strength, say, to complete a 170 Kg clean and jerk. The greater problem: still to toss the javelin, shoot skeet, triple jump, kayak, or dive, frenetically toggling between the same switches you’ve just used to provide a sprinter with speed in the 100 meters and to give a weight-lifter, at least theoretically, enough gas to complete a dead lift. The script for interaction shifts with every event and sometimes, even between training mode and trial/competition modes, leaving users in what promises to be the most immersive of interactive experiences—video games, after all, are direct descendants of the immersive arcade shoot-outs—pacing angrily through the slender instruction pamphlet, trying to figure out what functions the “A” and “B” buttons will signify for the next 10 minutes.

For our affective experience to remain immersive, both narrative and interface alike need to overtly guide or curtail our possibilities for action. Interactive games fulfill their promise as immersive when they offer us an obvious schema for narrative structure and interface, and when they offer us predictable, tightly scripted interactions, enabling us to enjoy virtual experiences that are either unattractively risky or denied to us in everyday life. In “Gadget,” if an informant in the Museum train station has instructions for you to turn your quest around, the train idles helpfully in the station until you venture out of the car, stroll up to one or two likely looking characters, and receive the vital clue. In the “Virtual Murder” series, Gilligan’s sidekick, a nameless help-mate portrayed initially by actress Sherilyn Fenn, is also an apt throwback to earlier narrative forms. While her cropping up in a police procedural is, perhaps, unremarkable, she remains the lone example in interactive games of the “Fifth Business” or the agent who exists solely to advance the narrative. In police procedurals, detective novels, and mysteries of all stripes, of course, the Fifth Business is usually the protagonist’s sidekick. Sherlock Holmes, famously, had Watson to bounce ideas off and to help him unravel clues, often unwittingly, and even Inspector Morse had the much-put-upon Sergeant Lewis. Both Watson and Lewis performed, like every good agent, the function of sniffing down false leads and interviewing suspects while their respective bosses got down to the real detective work and eventually solved the case—significantly, never without some intervention from their sidekicks. Of course, both agents also function as narrative foils to their bosses. Watson and Lewis are famously dim where their superiors are quick-witted, badly read and poorly mannered where Holmes and Morse are educated, cultured men of the world. Yet the agent is also the mystery’s unsung catalyst, a force who can usher the plot along efficiently precisely because he or she is an unobtrusive character never quite in the spotlight. Perhaps, given the paucity of scripts drawn from earlier media, the absence of agents from interactions should seem unremarkable. Yet the agent is such a potent tool, one that can clarify interface elements and possible actions, that we can only wonder why agents remain so strangely unused, apart from Microsoft’s brief, mid-90s foray into plugging an obnoxious agent named Bob into its desktop interface. An agent in “The Longest Journey” could make suggestions about what April Ryan ought to do with the
dinosaur scale and twig, saving you from scrabbling for the cheat-sheet walkthrough—decidedly an immersion-busting experience—similarly, an agent or voice-over could tell you how to position "Grim Fandango's" forklift in the elevator before you embark on your 40th attempt at halting the elevator or risk losing any remaining vestiges of sanity. Or you could enlist an agent or toggle the voice-over mode on during the early stages of the interactive, when you’re still determining the conventions and constraints governing the plot, characters’ actions, and environmental cues, only to leave both these guides behind once you’ve fully grasped the details and immersed yourself in the narrative.

PLEASURES OF IMMERSION, PLEASURES OF ENGAGEMENT

The pleasure of immersion in interactives stems from our ability to take guided action and see the outcomes from our choice of one or more scripts within a single schema. In contrast, the pleasure of engagement with hypertext fiction comes from users’ access to a wide repertoire of schemas and scripts, our attempts to discover congruences between the hypertext and an array of often mutually exclusive schemas, and, ultimately, our ability to make sense of the work as a whole. Even though Janet Murray’s list of plots as symbolic actions include sensemaking and assembling fragments into a coherent whole, Murray’s objection to what she calls “structured literary hypertext” reveals a criterion for aesthetic pleasure clearly founded only on immersion: “navigation unfolds a story that flows from our own meaningful choices.” Yet readers of modernist works like Mrs. Dalloway, The Good Soldier, In the Labyrinth, and Ulysses must actively wrestle with wandering narrative perspectives, tortuous representations of time, and deliberate disruptions in space, time, and causation, as well as the requirement that they ultimately understand the entire work relative to its spatial form. Anyone who confuses Great Works with an aesthetic of immersion should remember Joseph Frank’s famous declaration about Ulysses, which, he claimed “could not be read, only reread.” These texts engage readers deeply because they do not follow schemas for which readers can unhappily apply ready-made scripts. Instead, they violate existing conventions, switch schemas, and, in the case of works like In the Labyrinth, violate even our assumptions about continuity from one paragraph or even sentence to the next.

Yet even the earliest readers of the first published hypertext narrative, “afternoon,” experienced a kind of engagement that would have seemed familiar to readers of Ulysses or The Wasteland. First, readers bring to new media their schemas and scripts from older media, just as hypertext fiction itself draws on conventions inherited from print for plot, character, intention, and tropes. Second, published criticism which, in the cases of Joyce and Eliot, included symbolic “keys” to their work, helped readers in developing schemas to fit the new material. Joyce circumscribed notes that mapped Stephen’s and Leopold Bloom’s day in Dublin onto the adventures of Odysseus, while Eliot extensively footnoted his own poem. Readers of hypertext fiction, like Joyce’s and Eliot’s audiences, are more likely to seek out secondary sources to supplement their array of schemas for understanding the text. These “engaged” readers are also more likely to employ these schemas as simply part of a repertoire, rather than as sources for controlling scripts that determine singular interpretations of a work. Such secondary sources include Joyce’s own extensive criticism on hypertext aesthetics, as well as a growing body of criticism on the narrative significance of navigational mechanisms in hypertext. Third, even in a relatively new genre, some of its newly minted “grammar” and tropes are accessible to early audiences: cross-cutting, special effects, and the shot-reaction shot sequence all appeared during cinema’s nickelodeon era. In complex hypertexts, immanent structures include proximity signifying causal or relational connections between lexias in spatialized text, as well as recourse to remind readers of previously encountered lexias or to situate already encountered lexias in new contexts where they take on new meanings. Mark Bernstein has also identified cyclical repetition broken to signify closure, contour, where cycles coalesce or collide, and montage that establishes connections across the boundaries of nodes or links, as used by Landow. Jackson and Paul. Readers engaged with hypertext fictions like "Victory Garden" or "Twelve Blue" make hypotheses about the relationships between lexias and the significance of links, layering onto the print readers engagement with character, continuity, time, and space, further interpretations of the significance of spatial relationships and links between lexias, of link types and their conditions. Long-term engagement with the texts—the necessary rereading Michael Joyce describes—makes some relations immanent, nullifies some hypotheses, thwarts some navigational strategies, and generally enables readers to enlarge their repertoire of textual aesthetics still further. Finally, when hypertext episodes also represent causally linked lexias that generate narrative tension, readers may become immersed in the narrative. Even when immersion gives way to engagement, the immersive lexias or episodes can still act as a centrifugal force that compels us to become engaged with the narrative.
Immersion Into Engagement Into Flow

The “episode vortex,” as Jim Rosenberg notes, however, can just as easily frustrate readers, launching them into “foraging” for the next episode.9 While immersion may easily lure readers into interactive narratives and organize their initial engagement, replacing promised immersion with engagement can also frustrate readers, even when they can develop a script that situates their frustrated immersion as strictly intentional, a deliberate effect designed by the author.

Even in the throes of engagement, disorientation in hypertexts is potentially more disconcerting than the momentary discomforts we experience in other media, notwithstanding our budding repertoire of effects and gambits that signify. The dreaded “lost in hyperspace” problem is due partly to our awareness that hypertexts exist in virtual, 3D space—which may or may not be represented to readers via maps or spatial navigational tools—partly to our awareness that links often involve recursion and complex conditionals, seldom making visiting every lexia or link once the equivalent of experiencing the entire work. When we consider the affective dimension, however, the absence of guides for the length of time occupied by our engagement or immersion may be still more significant. When we sit down with a novel or settle ourselves into a Broadway theater or our local cinema, we know approximately how long our immersion or engagement will last. Book chapters, like film running times, often owe as much to the length of time readers rely on stages that signal reader’s progression through the text: “ages” for “Myst,” “realms” for “Obsidian.” Hypertext fictions, however, lack such clear signals to readers, making it difficult for readers to determine if their script-acquiring and developing have been successful in helping them understand the hypertext as a structure of narrative possibilities, or if they need to accommodate, modify, and generate still more scripts. Some writers have built forms of closure into hypertexts that enable readers to pause in their reading or leave it completely.26-34 But link conditions in complex hypertexts can yield different juxtapositions of lexia and fresh narrative possibilities, just as a familiar episode may branch in several unexpected directions the next time out, mitigating the cues potentially offered by these approximations of closure.

Finally, while immersion may shade into engagement — now an imminent development with recent calls in the interactive game industry for more backstory and narrative14 — and engagement into immersion, neither of these affective dimensions maps all that tidily onto most definitions of interaction.23 As Joyce26 and Aarseth23 have noted, readers of most hypertext fiction are merely exploring the narrative, not constructing its links and rearranging its structure, or even generating lexia and links themselves. While the advent of the World Wide Web and collaborative structures like Brown’s Hypertext Hotel32 suggest that hypermedia’s contribution to aesthetics may be a blurring of the line between reception and creation, the relatively limited interactions of immersive or engaging interactives should not likewise limit our quest for features, metaphors, and conventions that enhance our affective experiences.

Given the enhanced immersive possibilities of full-motion video, not to mention virtual reality, coupled with hypertext fiction’s complex possibilities for engagement, future interactives could easily enable casual readers to experience what Mihaly Csikszentmihalyi calls “flow,” a condition where self-consciousness disappears, perceptions of time become distorted, and concentration becomes so intense that the game or task at hand completely absorbs us.33 Since flow involves extending our skills to cope with challenges, a sense that we are performing both well and effortlessly, this state hovers on the continuum between immersion and engagement, drawing on the characteristics of both simultaneously. Prensky in the early 1980s, Sherry Turkle32 noted something like flow states in teenagers grappling with computer games. Where immersion involves identification with characters and narrative elements—the local details that keep us involved even when we know the plot’s trappings intimately—
engagement involves deciphering the author's or game designer's intentions. During a flow state, Turkle noticed, teenagers both identified utterly with the objects they were manipulating—the equivalent of "becoming the pinball," unthinkable to the player of the analog arcade game—and became deeply involved in determining the constraints built into the game. Most vitally, however, she noted that player after player was obliged to keep up with a rhythm dictated by the game itself, a "relentless... demand that all other time stop... and that players take full responsibility for every act." The combination of all three conditions, she realized when she interviewed inveterate game players from teenage social misfits and an ability to play strategically within its constraints.

Flow is, however, elusive, fleeting, and intensely problematic. The social misfits and uptight executives Turkle interviewed most likely achieved flow states during gameplay at least as much due to the all-but-non-existent characters or environment in the video games they played. Artists, writers, professional athletes, and musicians can experience flow states during practice or performance, as can connoisseurs of music, film, dance, or sports. For example, film critics may notice how deep focus, changes in film stock, and camera angles frame a sequence or allude to other films, an extratextual perspective on the film that is characteristic of engagement, even as they remain deeply immersed in the characters and plot developments of the narrative playing before them. Further, since engagement tends to focus our attention on the frame and materials themselves, texts like "Ulysses" or "afternoon" tend to immerse us only for short periods before demanding our engagement. As interactive, however, begin offering us worlds that increasingly resemble the one outside the text, and writers begin introducing them into their complex plots, characters, and orienting devices like voice-overs or agents, even casual readers may one day experience the flow that today only a privileged few enjoy when watching or creating narratives.

References

50

Electronic Art and Animation Catalog