TOWARDS A COMMON LANGUAGE: A THEORIZATION OF TEXT, AUDIO, AND IMAGE IN VIDEOGAME NARRATIVE

BY JUSTIN KEEVER

Senior Honors Thesis
University of North Carolina at Chapel Hill
Department of Communication Studies

(Advisor Signature)
# Table of Contents

Introduction – 1

I. Text-Primary – 7
   Ia. The Text Adventure – 9
   Ib. Developments of the Text-Primary Narrative in the Mainstream – 21
   Ic. The Text-Primary in the Modern Avant-Garde – 27

II. Aural-Primary – 32
   IIa. Audio-Log and Monologue – 37
   IIb. Travel Audio – 41
   IIc. Potentially Offscreen Bodies – 45

III. Image-Primary – 48
   IIIa. Embedded Narrative – 49
   IIIb. Cutscene – 57

Conclusion – 64

Works Cited – 69

Games Cited – 74

Filmography – 76
Introduction

The mention of the videogame medium in academic disciplines outside the realm of game studies tends (for this writer, anyway) to inspire excitement and dread in equal measure. The former would seem to derive from an appropriately childish desire to sit at the adult’s table of academia, as it were: long has there been discussion of what games have borrowed from the media that have come before, and little of the reverse. The latter feeling is based in the fear that the term “videogame” would appear, as it did for years, in a negative connotation. Not in the archaic media/sociological studies sense (i.e. the tired question of videogame violence and its correlation to aggression in players), but merely in the sense that the videogame medium is perceived either as a lesser art form, or perhaps as not an art form at all. It’s with these preconceived aspirations and worries I approached Thomas Elsaesser’s Chapter on the “Mind-Game Film” in his book *Puzzle Films: Complex Storytelling in Contemporary Cinema*, a text that was published in March of 2009. In this Chapter, Elsaesser cites a school of thought in modern film narrative analysis which posits that “mind-game films [are] leftovers of classical narrative, during a period of transition, when the default value of cinematic storytelling is rapidly becoming that of the interactive video-game and the computer simulation game” (22). Elsaesser continues by claiming that this assertion “should certainly leave the theoretician dissatisfied,” explaining that “the literature on whether games are narratives at all, or need to be seen as an entirely different species, is vast and vastly divided…” (22). Here, Elsaesser has provided game academics with a much-needed outsider’s perspective: Twelve years after Janet Murray wrote a book on the topic of interactive narratives, eight years after the supposed “Year One” of Game Studies (Aarseth), six years after Gonzalo Frasca claimed both that games are resolutely not narrative and that the debate between narratology and ludology didn’t actually occur (Simulation
versus Narrative; Ludologists Love Stories), and 5 years after Espen Aarseth’s “Genre Trouble,”
game studies still had not offered a definitive answer as to whether or not games can be
understood via narrative theory. This lack of resolution creates the image of a house divided: a
medium that persons from outside disciplines can only approach with a measured uncertainty at
best. This inconsistency obviously manifests itself in the myriad definitions of videogame:
compare, for example, Jesper Juul’s six formal features of the videogame – Rules; Variable,
Quantifiable Outcome; Values assigned to possible outcomes; Player Effort; Player attached to
outcome; and Negotiable Consequences – to Geoff Howland’s five elements: Graphics, Sound,
Interface, Gameplay, Story (Game Player World; 1998). Even the exact spelling of “videogame”
is left undecided: is it “video game,” as Mark JP Wolf refers to it in his book The Medium of the
Video Game? Or is it the “videogame” that James Newman and Grant Tavinor refer to in their
work? Or is it the medium of the “video-game” that Elsaesser speaks of? These kinds of
inconsistencies are not the growing-pains of a new academic discipline: commercial videogames
have existed for over forty years, and scholarship on videogames has been written since before
the supposed “Year One” of academic game studies thirteen years ago. This is a problem.

Worse, it’s a problem that is often ignored, or even celebrated. One need not look further
then Trammell and Sinnreich’s “Dynamic Ontology” of the game studies field, which, in its
attempt to “unite and distinguish the diverse connections, assumptions, and epistemic positions
among the various authors within the larger discursive field of game studies,” instead reveals the
absurdity of the variance found within a single discipline: “game studies” as defined by
Trammell and Sinnreich encapsulates eight theories and ten fields of study. This is not simply
diversity – this is the lack of a clear identity under the guise of diversity. This much is made clear
by the duo’s placement of Ian Bogost on their chart: the academic is placed in the theory of
“Media Studies,” despite spearheading the “Platform Studies” movement with Nick Montfort not 4 years prior. Indeed, the proliferation of so many courses of study (often by the same authors!) since Game Studies’ Year One speaks to a discipline-wide lack of focus, a group of academics searching blindly for any and all new perspectives and theories as though the answers for how to best approach the medium of the videogame lie anywhere but their current discipline. What’s even more disconcerting is that this aimlessness is sometimes better than the alternative: the prospect of a book entitled “How to Talk About Videogames” being written by the same man who once wrote that videogames would be better off without characters is nothing short of terrifying: while at the time of this writing the general public can only speculate about the contents of Bogost’s latest book, nothing in the man’s contemporary work would suggest any interest in a meaningful discussion of narrative, despite the fact that many (if not most) videogames have narrative aspirations.

This essay is meant to act as a part of the larger corrective to these damaging impulses of videogame studies, to treat videogames first and foremost as authored narrative texts, and to establish a useful critical vocabulary for approaching games in this manner. It is to this end that this essay has been fashioned as a narratological analysis of the videogame medium, specifically as an analysis of the manifestations of game narrative, both as they have existed in the past, and how those past manifestations exhibit themselves in contemporary videogames. To use the linguistic metaphor of traditional narratological studies, this paper is an examination of parole, the way in which the “narrative emerges from the narrative text” (Fludernik 8). However, in opposition to traditional narratological studies, this close analysis of narrative parole in the context of the videogame will not be limited to strictly to literal text (i.e. “how sentences turn

---

1 https://twitter.com/ibogost/status/576464559013232640
2 http://www.theatlantic.com/technology/archive/2015/03/video-games-are-better-without-characters/387556/
into narrative” as Fludernik suggests [8]): the videogame is a multimodal, composite medium, and thus provides a wealth of possibilities in terms of the manifestation of its narrative. For example, games like Zork or Planetfall work in entirely on the level of text, while games like System Shock 2 or its spiritual predecessor Bioshock communicate most of their narrative information via snippets of audio; and these different means of communicating each game’s authored narrative have profoundly different effects on the player’s experience of each game. It is the aim of this essay to assert the importance of these various creative decisions made by videogame developers: to explore how the varying modes by which developers choose to communicate their authored narratives affects the player’s experience of said narrative, and the time and space of the gameworld he/she is navigating. To that end, this essay will chronicle the lineage of the primary methods by which videogames communicate their authored narratives to players, methods that have been separated into three formal classifications: Text-Primary, Aural-Primary, and Image-Primary. It is important to note that the disaffiliation of these primary narrative methodologies does not presuppose mutual exclusivity: in practice, most modern games employ some combination of text, audio, and image in order to communicate a story. Thus the use of the word “primary” over a word like “only” in the naming of these categories: these genre titles assume the privileging of a single element, rather than its exclusive use. For example, a game whose narrative method is said to be “text-primary” is not a game that necessarily only utilizes text to communicate the authored narrative, but instead privileges its text over the elements that may appear on-screen during narrative interludes. This close examination of text-primary, aural-primary, and image-primary modes of game narrative will provide a greater understanding of the common narrative strategies that have emerged in the medium of the
videogame, and the way in which the strategies manipulate the medium’s formal elements of space and time in service of crafting meaningful experiences.

This essay is split into three chapters, one for each narrative genre listed above. The first chapter concerns the Text-Primary Narrative method is approached through the lens of literary scholar Mikhail Bakhtin’s theory of the chronotope, or “time-space.” This chapter begins by attempting to identify a Text-Primary chronotope via a close examination of the text-adventure *Planetfall*. As part of this exploration of what constitutes the Text-Primary chronotope, this paper will identify a structure of videogame time called player-action time, which will serve as a recurring concept for the duration of the essay. This analysis of the Text-Primary genre continues by examining the way in which the method was adapted into the mainstream after the text-adventure began to lose relevance: the narrative mode will be traced through early graphical adventure games and Western Role Playing Games. Given this essay’s strict focus on authored narrative in the videogame form multiplayer text environments like MUDs will be excluded from this conversation. Finally, the first chapter will address the way in which the Text-Primary Narrative emerges in the modern avant-garde, briefly exploring how recent art games have used this narrative mode to recontextualize and present a meta-commentary on the lasting influence of the text-adventure game.

The second chapter, which is focused on the Aural-Primary Narrative, is structured differently, beginning with an exploration of composer and theorist Michel Chion’s notion of the acousmatic voice in cinema. This essay will identify the differing ways in which videogames and film identify the source of a voice, and use this difference to explain how voices are granted more autonomy in the videogame medium, which allows for the development of audio-focused storytelling methods. From there, this chapter will explain three modes of aural-primary
narrative: Audio-log/Monologue, Travel Audio, and Potentially Offscreen Bodies, which are disparate in terms of their effect on the construction of the game-space, but parallel one another in terms of the player’s experience of time.

The third chapter will address the Image-Primary Narrative Method, which opens by placing the profound aversion to representation and aesthetics in videogame studies in the context of the pictorial turn that has taken place in recent criticism. This is a rhetorical move which is designed to allow the author to contextualize videogame studies’ rejection of the image in terms of a larger trend in academia, and expose this mistreatment of the image as simplistic and reactionary, thus allowing this essay to disregard this trend in game studies and explore the true importance of the image in the videogame medium. The third chapter explores two modes of the Image-Primary Narrative: the embedded narrative, a term inherited from Henry Jenkins that concerns the gameworld’s mise-en-scène, and the cutscene, which is a remediation of the medium of film in the context of the videogame.
I. Text-Primary

In 1937, Russian Scholar Mikhail Bakhtin adapted the term “chronotope” (time-space) from Einstein’s Theory of Relativity to the realm of literary criticism, citing its usefulness as the way in which “it expresses the inseparability of space and time” (84). Bakhtin identifies the literary and artistic chronotope as the “primary means for organizing time and space… a force giving body to an entire novel” – in essence, the chronotope is the set of functions by which time in a novel, or an individualized space within a novel, operates (250). For an example of how the chronotope is used in literary criticism, this turns to Bakhtin’s own analysis of the Greek Romance. According to Bakhtin, the generic chronotope of the Greek Romance is marked by adventure-time, or, a lack of “natural, everyday cyclicity,” the implication being that between the beginning and ending of the Greek Romance, there is no progression of biographical time. The adventurous events of the Greek Romance’s middle portion leave no evidence of physical nor psychological impact, and due to the lack of natural progression of time, those occurrence of these events is dictated entirely by chance or “fate” (Bakhtin 90-92). Furthermore, the “space” of the Greek Romance chronotope is that of the alien world: an un-exoticized, wholly foreign area that the protagonists of the novel have no pre-established connection to (Bakhtin 101). Ergo, the formula of the Greek Romance chronotope becomes “an alien world in adventure-time” (Bakhtin 102).

Bakhtin prescribes a focus on genre in regard to chronotopic analysis, saying that the chronotope defines “genre and genre distinctions” (85). In his essay, however, Bakhtin deals exclusively in analysis of novelistic genres (i.e. the Greek Romance, the adventure novel of everyday life, biography, etc.), which naturally privileges analysis of motifs, thematic content and plot structure over questions of technical form. After all, Bakhtin is writing about the
constructions of space-time in one very specific artistic form: the novel. Transferring the chronotope to the function of this essay requires a measure of recontextualization, beginning with an expansion of the term "genre." As Wolf notes, "video game genre study differs markedly from literary or film genre study due to the direct and active participation of the audience in the form of the surrogate player character." (114). The player-centricity of videogame genres has given way to genre classifications that operate almost entirely independently of the familiar elements like theme or iconography (two of the three areas where genre elements may emerge in American film, as identified by Ed Buscombe). Instead, genre distinctions are based entirely on the nature of the activity in which the player engages. Despite this departure from normal genre distinctions, it would be inaccurate to say that videogame genres are defined entirely outside the realm of iconography, as a certain level of visual representation is required to give the act of play the context necessary to facilitate genre classification: for example, moving a cursor and clicking takes on a widely different meaning in a First Person Shooter than it does in a point and click adventure game, and the differing contexts of the act of play are what delineates these distinctive genres.

All of this is to establish the precedent of a distinct function of the term 'genre' when applied to the medium of the videogame: specifically, a precedent of genres that are defined not by iconography and theme, but by mechanical form. It’s with this precedent in mind that I want to contextualize the three narrative methodologies this paper identifies: these are essentially broad genres of videogame, marked by their construction and narrative mechanics. It is with this understanding of these narrative methodologies as genres that this paper will adapt Bahktin’s notion of the novelistic chronotope, and apply it to the medium of the videogame in order to

---

3 The term “play” can be defined as an act 1) that is not obligatory 2) circumscribed within a certain space and time, 3) whose results are unknown 4) in which no goods or wealth are produced, 5) that is governed by rules 6) takes place in a second, make-believe reality. (Caillois 9-10).
identify the formal tendencies of text-primary, aural-primary, and image-primary videogames in terms of their respective constructions of spatiality and temporality. This application is made possible by acknowledging these narrative methodologies as a form of genre, albeit a conception of the term “genre” that refers to (narrative) mechanical elements rather than aesthetic or thematic commonalities. However, aesthetics will not go unexplored in the analyses of the three genres in question, as their unique methods of transmitting information would result in distinct visual styles.

Ia. The Text-Adventure

As the name implies, the “text-primary” genre of game narrative can be most identified by the privileging of text as the primary means of transmitting story information to the player. Therefore, the logical place to begin an exploration of this genre is with an early videogame form that would foreground text, and thus serve as the origin point of the text-primary game: the text adventure videogame. The text adventure videogame is perhaps the foremost “narrative” video game genre, insofar as it is the first genre whose act of “play” lacks the tactical or reflexive impulse of its agonistic or aleatory predecessors like Tennis for Two (Higinbotham, 1958) and Computer Space (Nutting Associates, 1971), and instead offering a focus on the consumption and comprehension of fictional prose; simply put, the text adventure is a genre built around reading comprehension, and as such, functions as the foundation for the “text-primary” genre of videogame story. Playing a text adventure game involves reading a description of an environment, or a transcript of dialogue, and reacting accordingly by inputting a command via a text-input line. Players must navigate their environment, collect and use objects, and speak to characters correctly in order to complete tasks with the overarching goal of completing the game; it is the navigation that most often takes precedence over these other sorts of tasks, as most of the challenge derived from the
text adventure involves being able to situate oneself within an space that is realized only through brief descriptions. As such, the narratives of text adventures can be understood in terms of Jenkin’s adaptation of De Certeau’s conception of narrative as “the transformation of place into space: the process by which “places,” which constitute stability, are acted upon by a narrative agent, who transforms those places into “spaces” by exploring and “colonizing” them (Fuller and Jenkins 66). Furthermore, the text adventure most often takes the form of a “tour,” a subjective journey through a space (Fuller and Jenkins).

The genre was first conceived with the creation of the game Colossal Cave Adventure, first developed by William Crowther in 1975, and expanded upon and promulgated by Don Woods in 1976 (Adams). The game was designed “at a time where text input (keyboard) and output (screen or printer) were common” indicating that the favoring of text descriptions over graphical depictions of environments was a practical choice, not an artistic one. A remake, titled Adventureland, was developed by Scott Adams and released in 1978 for the home computer. This remake would be promoted as a piece of interactive fiction, and would even be featured in The New York Times Book Review (Veugen and Querette 216). This rebranding of the text-adventure marks an acknowledgement of the narrative-driven impulse the Text Adventure would come to represent: the way in which the text adventure expunges graphical interfaces in favor of descriptive text allows for a greater capability – for the time – of environmental detail and more extensive dialogue, both key for developing more complex narratives (Holmes 28).

A text adventure that would be particularly remembered for taking advantage of these qualities is Planetfall, released by Infocom in 1983. Planetfall tells the story of a janitor stationed aboard a doomed interstellar spaceship; soon after the game begins, the ship begins to crash, forcing the player character to enter an escape pod and land on the planet below, where
he/she is tasked with finding a cure to a deadly virus that has killed off all life on the planet, lest the player character fall to the disease themselves. Of course, the player is not made aware of the ship’s ill fate immediately, and the game begins with the following description:

“Another routine day of drudgery aboard the Stellar Patrol Ship Feinstein. This morning’s assignment for a certain lowly Ensign Seventh Class: scrubbing the filthy metal deck at the port end of Level Nine. With your patrol-issue self-contained multi-purpose all-weather scrub-brush you shine the floor with a diligence born of the knowledge that at any moment dreaded Ensign First Class Blather, the bane of your shipboard existence, could appear.”

The player is given no objective that clearly indicates forward progress, as there is no action the player can take at this point that moves the story forward. The player must wait for the eventual explosion that signals the Feinstein’s demise which opens the escape pod doors, granting passage to the remainder of the videogame. In the meantime, the player can take incidental, largely meaningless actions, like scrubbing the floor, talking to the occasional passerby, or walking to a different level. Should the player elect to explore the Feinstein in the little time they have, they must input a command to walk in a compass direction (e.g. “walk east”), which will move the player to a new area. This means that each movement, so long as it as a movement that the game allows (e.g. you can’t walk north in an east-west hallway), results in the player moving to an entirely different space, and not a continuation of the space they were in before. Ergo, should the player decide to walk west while in the Feinstein’s Deck Nine, they will arrive in the Reactor Room, and not the westernmost part of Deck Nine.

To examine the chronotopal elements of the text adventure as they are present in *Planetfall*, this essay will begin by examining the way in which the act of “waiting” is made manifest within the game world. The incidental actions the player can take while waiting for the Feinstein to explode have no direct, diegetic effect on the Feinstein itself: the player-character can most accurately be called a bystander to this event, wholly removed from the explosion’s cause. However, it is inaccurate to say that the player does not cause the explosion, because the
explosion is triggered by the actions of the player: should the player scrub, walk, or speak enough times, the explosion event will finally trigger. Though players are essentially waiting for the game proper to begin, inactivity is not an option, as the plot will not move forward until the player takes enough actions. Interestingly, the player can type “wait” into the input line in order to make the game progress as well, though “waiting” won’t take up any more time than sweeping or moving around. The player could, if they wanted, simply type in “wait” over and over until the Feinstein explodes, which would correspond to complete inactivity on the part of the player-character. This incongruity between player activity and player-character inactivity reveals the way in which Planetfall, and text adventure games in general, operate temporally: the passage of time – and therefore the progression of a plot – are dictated entirely by player interaction. That is, the world and the narrative of the text adventure only react in response to player input, and never appear to act independently.

Mark Wolf identifies three manifestations of stillness in visual media: the photographic slide, the filmic freeze-frame, and the “still shot,” wherein a stationary camera is actively recording a scene with little to no perceptible movement (78). In the cases of these last two cinematic terms, there are indications of non-diegetic time passing in the form of the medium’s signature “grain, hiss, and flicker,” which indicate to the audience that the film is still progressing, despite the lack of diegetic movement. Because videogames use digitalized computer displays, there is no change in the color grain from frame to frame, meaning that these markers of non-diegetic time that are almost inherently present in film are not present in games. Instead, indicators of time passing in otherwise still scenes, Wolf argues, are introduced by the designers themselves: in games like Myst and Riven, for example, there will be “ambient sound of some kind,” sometimes accompanied by “ambient motion,” which takes the form of a rotating
fan blade or moving water. These movements will repeatedly return to their initial positions (or in the case of ambient sounds, the sound clips will repeat), which “combines stillness and movement in a form of cyclical or “looped” time.” (80).

It’s tempting to understand the progression of time, or the lack thereof, in the text adventure purely in the context of this notion of “looped” time. In a sense, looped time is at play in the text adventure: when waiting on player input, a cursor blinks repeatedly in the input line, a cyclical animation that signals to the player that the game is still running, and therefore, that time is progressing. However, the blinking cursor of the text adventure is not diegetic in the same way that the ambient motion and sound of Myst is; there is no blinking cursor in the world of Planetfall that awaits typed input from the player-character, making the blinking cursor more analogous to the flicker of film over a freeze-frame rather than a looped fan blade animation, despite the fact that the cursor is a deliberate design choice from the developer, rather than a medium-inherent feature. Due to the lack of a diegetic indicator of the passage of time, there is no reason for the player to conclude that time proceeds within the gameworld whilst the player remains inactive: just as characters in a novel do not linger in a living space when a novel has ceased to be read, the characters in a text adventure do not wait while time passes in the diegesis. Every pause in the text adventure marks a total pause in the progression of time, both in terms of the narrative and the gameworld; progression will only start again once a legitimate command is entered into the input line (phrases like “I don’t understand the word ‘use,’” which Planetfall displays when the player enters a command to “use” an object, also function as indicators of non-diegetic time passing).

While the text adventure does not necessarily utilize diegetic “looped time,” the concept is still drastically important to the text-primary genre of videogame narrative due to the way in
which looped time and the player reactivity that this essay has identified in the text adventure merge in later narrative videogames. Almost all games with graphic interfaces, which readily utilize looped time in the way Wolf describes, have also taken up a reactive model of narrative progression: that is, videogame stories proceed in response to the actions of the player. While the diegetic worlds of visual videogames don’t pause in the way the text adventure worlds do, the model of player input → plot output has remained more or less constant over the medium’s progression. For example, examine the way in which a sequence in the first episode of *The Wolf Among Us* (Telltale, 2013) parallels the opening waiting sequence of *Planetfall*: the main character of Telltale’s detective thriller walks into a bar, looking for a patron who is the prime suspect in a recent murder. The suspect is in the bathroom, a fact which neither the player nor the player-character have any way of knowing. In order for the suspect to leave the bathroom, the player must pass the time. However, the player can’t merely let the character stand in one spot and wait in real time for the suspect to finish in the bathroom – he/she must make the player-character perform a set number of menial actions, like talking to patrons or ordering a drink, in order to make the time actually progress in the game’s plot. The diegetic gameworld may operate on its own terms to some degree, but the story doesn’t proceed until the player takes action.

With this parallel in mind, it can be noted that the text adventure’s importance in terms of narrative progression of the videogame isn’t the repetitive use of the ludic “freeze-frame,” but the way in which the plot of the text adventure only progresses in response to player action, even when that action has no bearing on the events that occur next in the sequence of the plot. With this in mind, the serious work of defining the “time” element of the text-primary chronotope can be undertaken. Jesper Juul’s theory on game temporality becomes momentarily useful in this context, given that his simplification of the videogame as a “state machine” (a simplification I’d
normally consider reductive) is a useful way to understand the overt input/output functions of the
text adventure. Juul’s theory centers on the concept of *game time*, which consists of “the basic
duality of *play time* (the time the player takes to play) and *event time* (the time taken in the game
world)” (emphasis added). In action-focused real time games, these times are often aligned, and
the actions of the player are temporally matched with the corresponding actions of the player-
character; e.g. the way in which the player’s movements of the mouse correspond with the
movement of the player-character’s gun in *Quake III* (Juul, “Game Time”). However, Juul notes
that in other real time games like *SimCity*, the play time is not aligned 1:1 with the event time:
instead, in *SimCity*, the two minutes of play time required to build an in-game power plant
corresponds to roughly a year of event time – the time which it would take for a power plant to
be built in the real world. This allowance for disparity between play time and event time is where
Juul’s model becomes handy, as play time and event time in text-adventures is made almost
inherently disparate by genre’s imprecise sense of time. In *Planetfall*, play time consists of the
time the player takes to type in commands and the instantaneous results of those commands; it
takes a fraction of a second for the character to walk across a hall after the command “walk east”
has been input, for example (the textual descriptions in-game that indicate the character has
walked east would most likely [incorrectly!] fall under the label of “cutscene” in Juul’s model, as
they are non-interactive elements that interrupt the interactive play time. This facet of Juul’s
theory will be ignored, for reasons explained below). Event time consists of the imagined
diegetic time it would take for the player-character to walk down said hall. This time is measured
in-game via a “moves” counter (scrubbing a floor is 15 moves, waiting is 40, etc.) but the
imprecise nature of this measurement (just how much time is a “move”?) means that the event
time of *Planetfall* is indeterminate. Ergo, using Juul’s model of temporality, time in *Planetfall*
functions in such a way that the seconds it takes to type in a command and see an output in play
time correspond to an indeterminate event time.

Of course, Juul is not interested in the progression of game narratives, but in game states,
so this observation leaves the question of the text-primary chronotope unresolved. In order to
apply this model of temporality to narrative, the understanding of “output” needs to be changed.
Due to the simple input/output nature of the text adventure, and the imprecise nature by which
time is measured in those outputs, one can posit that each output in *Planetfall* constitutes a
narrative sequence: that is, every output in a text adventure corresponds to meaningful story
progression, as each move involves the exploration and familiarization with a place, thus
transforming it into a “space,” as outlined in Jenkins’ notion of the spatial narrative.

Additionally, Juul’s concept of what constitutes “play” in terms of game temporality can safely
be expanded upon in this context; Juul’s assertion that “if you cannot influence the game state in
any way… you are not playing a game” discounts time spent absorbing narrative information,
i.e. watching cutscenes and reading descriptions in text adventures, which is integral to the
experience of “playing” narrative games (“Game Time”). Ergo, the time the player spends
reading the text in *Planetfall*, as well as the inactive time he/she spends mentally solving an in-
game puzzle will be included in this discussion’s definition of “play time.”

Juul’s notion of game temporality has now been adapted to the player input → plot
output structure of the text-primary game. With play time (i.e. player input) now including player
inactivity whilst playing a game, the function of the blinking cursor of *Planetfall*’s input line
must be reexamined in terms of Wolf’s notion of looped time. Wolf does little to establish the
function of looped time in a narrative context, focusing instead on the way in which looped time
functions as a means for the player to learn patterns and thus master the game’s challenge, or as a
result of the player failing and being forced to re-experience a sequence of in-game events (81).

The concept of looped time is useful beyond these applications: the example of the fan that Wolf highlights in *Myst* and *Riven* is a gesture toward the way in which looped time functions as a kind of incidental purgatory in narrative games, a means by which the time of the game world is held at a standstill while waiting on player progression, all whilst simulating autonomy. Thus, the freeze-frame that accompanies the blinking cursor in *Planetfall* functions in the same way as loop of a fan blade in *Myst*: a means of affecting the gameworld in such a way as to make the progression of the narrative seem natural despite the potential inactivity of the player.

Given this paper’s interest in narrative, a question remains: what effect does this looped time have on the characters within the game’s narrative? To answer this question, one can turn to Bakhtin’s aforesaid notion of adventure-time. Bakhtin identifies adventure-time in his analysis of the Greek Romance, noting the way in which the time of the “improbable number of adventures” the characters endure is “not measured off in the novel and does not add up.” Significantly, “this time… is not registered in the slightest way in the age of the heroes” (90). The Greek Romance is marked by two biological moments in the lives of the characters – the arousal of passion at the beginning of the novel, and its “satisfaction” at the end (90). Bakhtin’s application of “adventure-time” as referring to the overall plot structure of a literary work isn’t particularly useful in defining the time of the text-primary videogame, but the notion of unmeasured time in-between two moments of biological development can easily be adapted to this paper’s adaptation of looped time: adventure-time functions in the text-primary game insofar as the looped time that occurs during play time in between plot sequences (the player input that precedes plot output) has no effect on the progression of any character’s arc, nor any plot-relevant progression of time in the gameworld overall. This kind of time will be defined by this paper as *player-action time*. 
Player-action time is marked by a structure of looped-time sequences of play with no biological significance, interrupted by plot-progression relevant sequences which have a meaningful biological importance in the development of the game’s characters. This player-action time constitutes one half of the text-primary chronotope, the “time” element of the “time-space.”

The space of the text-primary chronotope is both more abstract, and more easily identifiable. But, before the space can be defined, the way in which space operates in the videogame medium must be briefly examined. Mark Wolf notes the ways in which videogame space differs from that of the visual media that preceded it: First, there is no pro-filmic referent in the construction of videogame space, so “not only the representation of space, but even its implication, depend on being programmed and actively created, meaning that space could be structured in a different manner from other visual media. Additionally, the videogame “often gives the player some control over the point of view, allowing one to choose which spaces appear on-screen or off” (Wolf 52). From there, Wolf names eleven categories of videogame space, which are differentiated from one another based on the organization of on and off-screen space. Espen Aarseth notes the way in which “computer games are both representations of space (a formal system of relations) and representational spaces (symbolic imagery with a primarily aesthetic purpose)” (“Allegories of Space” 163). Aarseth expands on this understanding of videogame space to understand spatial representation as a “reductive operation leading to a representation of space that is not in itself spatial, but symbolic and rule-based,” using these “reductions as a means to achieve the object of gameplay” (“Allegories of Space” 163). In the service of this interest, Aarseth identifies a series of “distinctions” of spatial representation: first, he recognizes spaces that serve “man against the environment” approaches, wherein there is a clear distinction between the player-character and the world they inhabit (“Allegories of Space”
Aarseth also identifies a distinction between two spatial representations: “indoors” versus “outdoors,” where indoors refers to labyrinthine, restrictive environments, and outdoors refers to open landscapes (“Allegories of Space” 159). A videogame space can symbolize an outdoor area, but can be sufficiently restrictive as to be considered an “indoors” spatial representation: the outdoor environments of Myst are an example of this (Newman, Videogames 114). Finally, Aarseth identifies a third distinction of spatiality, based on the level of control the player has over the shape and form of the gameworld itself: simulation games like SimCity offer a high level of control, whereas 3D action games like Doom offer no meaningful influence over their environments (“Allegories of Space” 159).

These prevailing theories of game spatiality provide a solid foundation from which to identify the “space” of the text-primary chronotope. Text adventures such as Planetfall fall under Wolf’s first category of videogame spatial structures: “no visual space; all text based” (53). Wolf submits that the way in which text-adventures describe scenery without the use of images precludes text adventures from having on-screen space; instead, the entire game consists only of off-screen space. Alternatively, Wolf submits that “that which is currently being described by the text on-screen” constitutes the on-screen space – a statement which Wolf himself seems dubious of (54). It is the opinion of this author that the label “no visual space” is somewhat reductive: the core gameplay of the text adventure is, after all, navigation through (and mastery of) a puzzling environment, which would effectively align the text adventure with the “man versus the environment” distinction that Aarseth defined. In fact, Aarseth specifically cites the “textual descriptions in early adventure games” as a means of creating a “personal perspective on the gameworld” in the man versus environment spatiality (“Allegories of Space” 159). It’s this personal perspective that allows the text adventure to serve as the spatial “tour” narratives that
Jenkins identifies – a narrative form that is based around experience in and mastery of an environment. Specifically, playing a text adventure involves the mastery of an “indoor” environment: the restrictive nature of the compass-direction walk commands commonly used in text adventures ensures that environments are designed as labyrinths. For example, the way in which the input “walk east” in *Planetfall* takes the player to an entirely different area from where they were last (as long as there is a space to the east) epitomizes the restrictive, indoor nature of the text adventure: space is organized as a series of rooms, rather than an open expanse, even when the textual description of the environment would indicate otherwise. Ergo, the spatial structure of the text adventure can be understood as an “indoor” environment organized into rooms, presented through a personal perspective. However, this description does not cover the entirety of the text adventure’s spatial structure: as reductive as Wolf’s subcategory is, the lack of visual space does create a more unique spatiality in textual videogames: specifically, a tension arises wherein the organization of each “room” of the text adventure is fixed, creating a concrete organization of a place, but the space in and between these rooms is abstracted. *Planetfall* contains descriptions of relatively larger rooms and longer hallways, but these areas can be traversed via a single input, just as with any other space. Thus, diegetic space in *Planetfall* operates in a similar manner as diegetic time: it is largely indeterminate. The text adventure is uninterested in the exactitude of the dimensions of its space – its focus lies in creating an environment of rooms with consistent, logical connections to one another. The Rec Room is to the east of Dorm A, and Dorm A is to the west of Rec Room. With this in mind, the space of the text-primary chronotope can be understood as an indoor environment organized as a series of individual rooms, with a concrete position relative to one another but indeterminate dimensions. Because of the way in which text adventures function as tour narratives, it can also be assumed
that the space of the text-primary genre is presented from a personal perspective, and that the space is, in all actuality, an alien “place” that will be transformed into a “space” via exploration. Hereafter, this kind of space will be referred to as the “textual labyrinth.”

The text-primary narrative chronotope can be defined as “player-action time + textual labyrinth,” an equation which reflects the complex way in which the text-primary videogame abstracts time and space in order to provide a narrative experience that accounts for player inactivity and presents a logical, structured gameworld that isn’t muddled by the desire for exactitude of dimensions. The formulation of this chronotope has been entirely based on an exploration of the text adventure, the foundational genre of the text-primary narrative, and a genre that has fallen out of favor. 1983, the year of Planetfall’s release, would be the final profitable year for Infocom, who would be dissolved by Activision in 1989 (Holmes 36). However, this chronotope would remain relevant, as new games would adapt this chronotope in new contexts, further developing the text-primary narrative genre.

Ib. Developments of the Text-Primary Narrative in the Mainstream

Volition’s comedy-action game Saint’s Row: The Third (2011) features a mission named “http://deckers.die” in which the player-character, the head of a criminal organization ironically called “The Saints” is transported into a Tron-esque cyberspace in order to carry out an attack on a group of hackers called “The Deckers.” This attack is, for the most part, carried out in the way the player-character confronts all of his foes: gun violence. For the majority of the level, the player is tasked with shooting the digitized Decker minions who stand between you and the boss character at the end of the level; that is, until the player reaches the first objective marker. Here, the screen quickly fades to black, and a brief block of text appears on screen, reading:

“You awaken to find yourself on cold flagstone. As you stand up you notice that the small room is lit by two torches. To the north there is a door, to the south a hole in the wall. What do you do?
Open the door
Take a torch
Go through the hole
Take a nap”

As the player scrolls through these options, the tapping of a keyboard can be heard. The voice of a secondary character can be heard off-screen, saying to the player character: “Ooo, I love these!”

This fun little meta-textual reference to genres past unintentionally serves as an important indicator of the way in which the impulses of the text adventure have been adapted into newer forms of videogame. The player must choose one of the four options listed beneath the question “What do you do?” Should the player choose one of the wrong options, the result is instant death, forcing the player to restart this brief sequence. Thus, what *Saint’s Row the Third* is doing in this sequence is not just imitating the aesthetic of the text adventure videogame, but also its basic gameplay structure: submit the correct input or die. What’s particularly important, however, is not the way in which *Saint’s Row* imitates the structure of the text adventure, but how it alters and streamlines it: the gameplay possibilities are made explicit to the player, deemphasizing the trial-and-error experimentation of games like *Planetfall* in favor of evoking a sense of immediacy and honesty about the limited possibilities of the play-space. Volition is essentially offering a visual interpretation of the inherent limitations of early text adventures: the wide range of actions promised by the ability to input any word or phrase in the input line of the text adventure is undercut by the limited number of responses the game can actually identify. In *Planetfall*, the player can easily be left fumbling in the dark, desperately attempting to find the right verb to use in combination with a certain in-game object, only to receive “I don’t know what x means” in response. By making the player’s situational options explicit, Volition has created a modernized vision of the text adventure, in which the player is cogent to all of the
possible actions they can take, accelerating the pace of progression while still maintaining the
genre’s basic chronotope: this sequence operates on player-action time, and the play involves
navigating a (simple) textual labyrinth. Therefore, this sequence exemplifies the way in which
the core generic elements of the text adventure can be preserved, even as the form is adapted.

Of course, *Saint’s Row the Third* is not the first game to adapt the text-adventure’s
genetic structure, but the way in which it directly cites the text adventure by mirroring its
aesthetic functions as a helpful indicator of exactly how the text-primary narrative genre emerges
in non-text adventure videogames. The implicit argument to be drawn from this sequence is that
as graphic interfaces became more common, and the text-input gameplay of the traditional text
adventure fell out of favor, option-based gameplay became the domain of the text-primary genre.
Now, “option-based gameplay,” in the context of this argument, refers specifically gameplay
wherein a series of choices are laid before the player in a text-based, menu-like format, and the
player is required to pick one of these given options to proceed (without this distinction, one
could argue, a là Sid Meier, that all games feature option-based gameplay [Rollings and Morris
38]).

Of course, other graphical videogames integrate option-based gameplay more elegantly
than *Saints Row* – in fact, it is easiest to turn to the next iteration of the adventure game, the
point-and-click adventure videogame, to comprehend the way in which the text-primary
narrative genre evolves. The assimilation of the text-primary and image-primary genres (the
latter of which will be expounded upon in a later chapter) is evident in *The Secret of Monkey
Island* (Lucasfilm Games 1990), wherein the game screen is separated along a horizontal plane
into a graphical depiction of a game environment and a textual list of actions the player can take
at any time by clicking on one of those options, then clicking on the corresponding item within
the graphical environment the player wanted to interact with. With this structure, it is immediately evident how the core gameplay loop of the text adventure is preserved: the basic “verb/object” logic of the text input system is constant, despite the introduction of the graphic interface. However, the preservation of this gameplay construct does not indicate that *The Secret of Monkey Island* belongs to the text-primary genre: the player’s understanding of the world he/she navigates is situated in its visual representation, not in the text that dictates how the player can interact with said world. However, that very emphasis shifts away from the world as it is graphically represented during the game’s numerous conversations: in these sequences, the normal process of play is literally stripped away, as the menu of potential gameplay actions is replaced with a list of potential lines of dialogue that the character can utter. This kind of interactive dialogue sequence has become commonplace within narrative videogames; the general experience of which is rather succinctly described by a group of Bioware employees in their patent of their own unique graphical interface for interactive videogame conversations:

“Many different types of videogames have been produced, which enable the player to take control of a character in a game situation that involves a storyline. The storyline may progress using simulated conversations. In a role-playing game, the conversations may be held between the player's own character and one or more characters in the game. Using prior art interfaces, typically the conversation progresses through the player making choices between different dialog responses presented as lists of text on-screen. Selection of different responses results in different paths through the conversation or story. Making a dialog choice results in the character “speaking” the chosen line. Some times such videogame dialog is implemented by displaying the character's dialog line in the form of text, and at other times it is implemented through playback of pre-recorded lines spoken by a voice actor. In either case, when players are presented with a dialog choice, they are usually required to refer to a text-based description or another set of indicators to understand the available choices” (Hudson et. al).

The final line of this description is particularly significant in this essay’s context: interactive conversations, even in graphically oriented videogames, re-orient the player’s focus towards a textual element, privileging text as the primary means by which story information is communicated to the player.
This re-privileging of text in interactive conversations allows these sequences to serve as something of a microcosmic iteration of Huizinga’s notion of the magic circle within the videogame itself: the interactive conversation often provides a sufficiently different play dynamic, a temporary world within the ordinary world of play, “dedicated to the performance of an act apart” (Huizinga 10). As much is evident when examining *The Secret of Monkey Island*, as the way in which the normal UI elements are stripped away in favor of the dialogue options of when an interactive conversation is initiated gestures towards this dynamic: the normal gameplay of spatial navigation and puzzle solving is removed entirely, replaced with a static situation that requires little to no spatial awareness beyond what has already been established when the conversation is first initiated. The focus of the game shifts away from the overarching problem of narrative progression via environmental puzzle solving, offering instead an independent scenario that the player must navigate. In the case of interactive conversations, the “textual labyrinth” of the text-primary chronotope is preserved, albeit metaphorically: differing dialogue options function as different paths for the player to take through a branching map of authored speech. This map (often addressed via the metaphorical term “dialogue tree”) provides the player with a fully textual environment to navigate, which can operate within the context of a graphic-focused game while remaining functionally independent – a performance inside of a performance.

This basic structure of interactive conversation would remain more or less constant as the videogame continued to evolve graphically: the playable dialogue sequences in Western RPGs like Bioware’s *Mass Effect* (2007) operate in essentially the same manner as conversations in *The Secret of Monkey Island*. The demanding shooter/exploration gameplay of Bioware’s space-faring RPG is set aside when conversations are initiated; what’s more, the tedious shot/reverse shot most of the game’s dialogue sequences are framed in almost serves to de-emphasize the
cinematic nature of the game, pushing the player’s attention towards the dialogue wheel that sits at the bottom of the screen during these segments, or the subtitles the player may have opted into. Regardless of the sudden introduction of Hollywood continuity editing and “cinematic” framing (features that will be more readily identifiable in the image-primary narrative genre), *Mass Effect*’s conversation sequences provide are prototypical examples of text-primary narrative: the dialogue tree is maintained (and emphasized via outward-branching dialogue wheel graphic on the screen’s lower half), and, importantly, the conversations occur at player-action time: the time between the player being prompted to respond to an NPC and the actual input of said response does not take up any kind of biographical time, insofar as these intermittent pauses in conversation are never acknowledged by any of the characters, and conversation will not resume until the player has responded.

While the domain of the text-primary narrative genre in most modern videogames is interactive conversations, it’s worth noting that the text at work in these sequences isn’t always limited to dialogue. Often, the grand scale of role playing games like *Mass Effect* requires that the development team reuse certain generic animations, or have certain contextual actions occur offscreen. In these cases, the text functions as stage directions, giving clear context to actions that have been abstracted in the game’s visual presentation via either a synchronous description or a preceding option that corresponds to said action. These kinds of descriptions are rooted in western Role Playing Videogames of the mid-to-late 90s, whose prevalent textual descriptions provided concrete details of character’s actions in a manner similar to later role playing games, albeit with a decidedly more literary style of prose. These early graphic role playing games were mostly presented via an isometric, god’s eye perspective with crudely animated characters. This meant that the actions of the characters, their appearances, and the intricacies of the world itself
required textual elaboration. Role Playing Videogames like *Planescape Torment* would borrow the divided screen of *The Secret of Monkey Island*, which, in turn with the more prolix descriptions, creates an even stronger focus on the textual storytelling. The descriptions of games like *Planescape Torment* read like meta-ekphrasis, in which the artwork being described is the game itself. The textual descriptions become the primary means by which the player comprehends the world in which they are placed. The textual labyrinth of the text-primary chronotope is preserved.

**Ic. The Text-Primary in the Modern Avant-Garde**

As important as it is to understand the way in which the text-primary narrative genre would develop and be adopted by the mainstream, it is equally important to acknowledge how independent art games reflexively acknowledge and/or subvert the standard construction of a text-primary narrative. In doing so, we can observe how game makers at the fringes of the medium have responded to the practices and assumptions the underlie the text-primary narrative as it is described in this essay, and understand how those assumptions inform the creation of these games, in spite of the way the specifics of the text-primary genre have gone undefined until now. Essentially, the text-primary narrative provides the savvy player with an easily identifiable ethos on which a select number of modern independent or avant-garde games are based, and thus provides a deeper understanding of the way in which these games both derive from an identifiable form whilst destabilizing and/or advancing that same form.

Take, for example, Christine Love’s 2012 game *Analogue: A Hate Story*, a game in which you take on the role of an engineer hired to download the log files of a wrecked spaceship. The game briefly informs the player of the context via a *Star Wars*-esque text scrawl (which is written like an email addressed to the unnamed player character), then presents the player with a
fictionalized DOS screen that serves as the “override terminal” of the spaceship computer where the desired files are located. The player is presented with a list of available commands:
“download, disable_ai, enable_ai, help, su, quit.” A cursor blinks below this list of commands, awaiting player input. Each game that this paper has cited, up to this point, that has made its affordances explicit has also allowed the player to simply pick one of the gameplay options by simply selecting it. This is not the case in Analogue; here, the player is required to type out these commands manually, à la Planetfall. The challenge of this early sequence does not simply derive from picking the right option, but also from understanding the grammar of the override terminal. Progression is not as simple as typing “download” or “enable_ai”: the verb/object logic of the text adventure is utilized in Love’s game (should the player wish to enable an AI, they must also choose which program they wish to activate), with the addendum of understanding how that command should be structured. Essentially, the game is tasking the player with deciphering a deceptively simple, fictional computer code. Furthermore, although Analogue: A Hate Story ostensibly takes place hundreds of years into the future, the game never makes the effort to present the player with a concrete representation of its diegetic world: instead, the diegesis is limited to the computer terminal that the player interfaces with. As such, Analogue takes the DOS screen aesthetic and text input gameplay of the text adventure and recontextualizes it in such a way as to call attention to the way in which the player is inherently interfacing with a computer while playing a game like Planetfall. The construct that the player must navigate in order to progress is the computer itself, a feeling heightened by the clean user interface that appears when an AI is activated, and the gameplay shifts to a focus on navigating through log files: the “textual labyrinth” for the player to explore becomes, in some ways, more literal, insofar as there is no physical space that the textual descriptions represent, and the labyrinth
becomes the arrangement of the text itself. This sort of play has an interesting effect on the
game’s construction of time as well: because there is no differentiation between the actions the
player performs and the corresponding actions of the player character, the play time/event time
relation, to once again use Juul’s model of game time, is 1:1. However, the narrative still has the
appearance of progressing via player-action time, as the actual progression of the game’s story is
based in the behavior of the player, with no acknowledgement of any extreme passage of time on
the part of the AI characters, even though that passage of time is no longer cyclical. But this
change in the nature of game time gestures towards Analogue’s reflexive awareness of the kinds
of play it is drawing on, and the truth of what it is we do as players when we play a text
adventure, or any videogame.

Other games invert the models of games like Planescape Torment, using text as a means
of abstraction rather than fabricating a concrete sense of place. The game Nier (Cavia, 2012)
features a sequence in which the player character must interact with the residents of a small
forest village, who have all been put into a trance-like state by a nebulous force. When the
unnamed player-character speaks with one of these villagers, he enters the same trance, which is
indicated by the screen gradually fading to black and descriptive prose filling the frame. These
sequences are structured similarly to the aforementioned “text adventure” sequence in Saint’s
Row: The Third, in that the player will be given a series of options in response to the descriptive
text they just read, and must choose the correct option (or series of options) to proceed. These
options include answers to questions posed by NPCs in the textual trance, and navigational
options (eg walk north, walk east, etc.). While the incongruence of the “text adventure” sequence
in Saint’s Row: The Third was played for comedy, the lack of continuity between the graphically
represented world of Nier and the textually depicted trance-world has a deliberately surreal
effect: the slow fade to black (as opposed to *Saint’s Row*’s wipe transition) creates the sensation of a slowly loosening grasp on reality, leading to a moment in which these two diametrically opposed realities are superimposed onto one another, creating a vivid sense of uncomfortable dissonance. Not only does the textual description of the trance-world contradict the graphical (there will never be a graphical-world referent in the textual world), the textual descriptions are free from the aesthetic tropes of text adventures, confronting the player with an abstractly defined sense of place that borders on incomprehensibility. In essence, *Nier* pulls the player from the comfort of the graphical world by contradicting its own reality with a textually depicted one, confronting the player with the differences of the experience graphical and textual worlds: the latter, which seemed so natural in the videogame’s developmental years, enters the realm of the surreal.

*Limits and Demonstrations* (Cardboard Computer, 2013) has a moment similar to *Nier*’s trance sequence, although with a different aim in terms of its deconstruction of the text-primary formula. This short game features a small group of people attending an exhibition of a fictional artist’s works, all of which are highly ambitious and resolutely avant-garde. Most of these works are visual, save for one, which consists of audio tape of the artist and a group of companions exploring an office building. As the player-character listens to the tape (the content of which is displayed via on-screen text, as there is no voice acting in *Limits and Demonstrations*), the screen fades to black, and it is revealed that the tape is structured like a modernized text adventure, prompting the “listener” with a series of options at certain points in the recording. These options are given via very precise instructions, such as “to leave room, rotate seventeen degrees and advance four inches.” It soon becomes clear that these instructions are not referring to the movement of the recorded people within the space, but to what the listeners must do with
the tape in order to select an option within the art exhibit itself. In doing so *Limits and Demonstrations* confounds the typical notion of the textual labyrinth, because the player-character’s actions, which the player directly controls, do not correspond to movement within the space that the player character is attempting to navigate. Progression is not tied to navigating a fictional space, but a fictional artifice that corresponds to a fictionalized space: in essence, the player is controlling a character who is also acting as a player, directing the movements of the character as they play a game. The player is still navigating a textual labyrinth of sorts, as the audio-tape artifice is presented almost entirely via text, but the navigation is occurring with a degree of separation that complicates (but does not disprove!) the chronotope of the text-primary narrative.
II. Aural Primary

French composer and academic Michel Chion has written that there is an inherent dualism in sound cinema. Due to the physical nature of film, the medium inherently makes an incision between the body and the voice, before attempting to restitch the two together at the seam (Voice in Cinema 125). In film, Chion states, bodies and voices are inscribed onto separate surfaces (the celluloid image and the soundtrack, respectfully), a process which constitutes this inherent dualism: only by the process of synchronizing the images and the soundtrack can cinema create the illusion that the voice and the body are being presented as one singular entity (Voice in Cinema 126-128). It is via this process of synchronization, Chion posits, that voices are grafted onto “a particular body that is assigned symbolically to the voice as its source” (Voice in Cinema 126). This symbolic anchoring of the voice is achieved by showing a mouth: if the filmic voice matches with the movements of a character’s mouth onscreen, than the audience can safely attribute that voice to that visible body – and if there is no visible mouth, than that voice is “offscreen” (Voice in Cinema 127).

To an extent, this audio/visual duality has been inherited by the videogame: voice performances tend to be recorded independently of the visual rendering of scenes, and even if these processes overlap (as once could argue in the case of motion/performance capture), visual and audio data are recorded via separate devices and stored in different files within the game itself. In this way, the videogame is invested in the same “restitching” process as cinema: rendered bodies and recorded voices are presented simultaneously in order to create the illusion of a unified being. It stands to reason, then, that games are invested in a similar process of synchronization: the juxtaposition of audio and visuals that creates a “temporal co-incidence… as a sort of guarantee that we’re in the real world” (Chion, Voice in Cinema 128). However, this
1:1 adaptation of our understanding of sound and synchronicity from film to videogames is an imperfect one, because as voice performances became more prevalent in games in the late 90s, the meagre graphical capability of the console era meant that the medium lacked the ability to represent the human body with the sort of fidelity that facilitates the words-to-lips synchronization of film.

Hideo Kojima’s *Metal Gear Solid* (1998) provides a useful example of this issue. As series protagonist Solid Snake enters the Russian compound that comprises the remainder of the game, players are shown a cutscene that depicts Snake hiding behind cover and calling his superior on the radio. In this shot (fig. 1), players see, for the first time, the in-game Solid Snake’s face (as opposed to a stylized facsimile that will be addressed later) as he is speaking. This moment of de-acousmatization is complicated by the three-dimensional model of the character himself: Snake’s sparsely detailed face doesn’t allow for any movement of the mouth, meaning that the *Metal Gear Solid* is incapable of resolving the source of Snake’s voice in the same way as a film.

![Fig. 1. Solid Snake infiltrates Shadow Moses](image)

4 Metal Gear Solid, along with Final Fantasy VII – which was released a year prior – are credited as the beginning of the “cinematic” videogame (i.e. a game with a heavy usage of “cutscenes:” typically non-interactive visual sequences that imitate to a certain extent the style and framing of film). However, Final Fantasy VII forgoes voice performances, rendering it irrelevant in the context of this essay.
Given that the mouth-to-word synchronization that Chion identifies as the suture of voice and image in film couldn’t be accomplished in the primitive days of the videogame, the medium is forced to develop without a key capability of representation in audio-visual media. Videogames approach this problem in two primary ways, the first being the appropriation and acceptance of acousmatic sound as a core storytelling conceit, using voice-over narration and diegetic acousmatic instruments (i.e radio, telephone, etc. [Chion, *Audio-Vision* 71]). The second way the medium addressed this issue was an expansion of the symbolic source of voices: in videogames, the body itself becomes the symbolic source of the voice, rather than the mouth.

These methods of resolving acousmatic sound require further exploration, both in terms of how this re-definition of the source of the voice functions on a practical level, and how it is indicative of a certain “looseness” of videogame speech that informs how games tell stories aurally. In *Metal Gear Solid*, Hideo Kojima “sutures” (to use Chion’s term) voices to bodies by means of two methods that still inform the relationship between voice and body in the medium today: via cinematic framing and synchronous bodily movement. Another early cutscene in *MGS* provides examples of both: Solid Snake hides in an air duct, secretly watching as two guards converse. The guards are shown in long shot, one near the center of the frame, and the other on the far right, their motionless faces covered by balaclavas (fig 2). Their concealed faces and low-quality rendering make it difficult to immediately ascertain which of the two guards is speaking, but the framing of the conversation suggests that the first belongs to the leftmost guard, as he is blocked in such a way as to be the more prominent figure in the frame. This suggestion is confirmed when the second guard begins to speak, as the camera pans to the right, reframing the conversation so that the rightmost guard is positioned in the center of the shot and the leftmost guard is removed from the frame almost entirely (fig 3). By adjusting the perspective, Kojima is
able to implicitly identify the source of the voices without explicitly showing the movement of the character’s mouths. Furthermore, when the in-game cinematic cuts back to Solid Snake, his head visibly moves as he speaks, even though the features of his face remain completely still (figs. 4 & 5). With this motion, the basic logic of audio/visual synchronicity is maintained, insofar as onscreen movement denotes a voice’s source: it is only the exact nature of the movement that has changed. Snake’s movement of his entire body expands the symbolic source of the voice outward from the lips. The resulting suture is less accurate than that which is typical of film, as this bodily movement ignores the complex process of phonation that is suggested by the movement of lips. As a result, Snake’s voice appears to simply emanate from his body.

![Fig. 2 (top left) and Fig. 3 (top right). Two guards speak to one another. These two angles are unbroken by a cut](image1)

![Fig. 4 (bottom left) and Fig 5 (bottom right). Solid Snake moves his head whilst reacting to the guards’ conversation](image2)

Even as graphical technology improved, these methods of de-acousmatization have endured: accurate lip synching isn’t needed for players to comprehend the source of videogame voices. The recently released supernatural teen drama *Life is Strange* (DontNod Entertainment,
functions as a useful case study in this regard: despite a generally positive critical reception, mainstream outlets have commented on the exceptionally poor quality of the lip synching (journalist Patrick Klepek quite comically expressed his frustration by simply writing “Those mouths aren't really saying the [subtitles] shown at the bottom of the screen, my friends!”). Despite this shortcoming, *Life is Strange* is able to successfully stitch voices to their sources because of the way in which videogames were forced to relocate the symbolic source of voice. The game extensively utilizes cinematic shot/countershot sequences in order to position speaking characters in the center of the frame, often asserting their status as the source of whatever voice is currently being heard (figs. 6 & 7). The game’s lip movements, while failing to serve a filmic de-audiovisualizing function, serve that function in the context of the videogame: while the lip motions are asynchronous with the exact lines being spoken, they are still synchronized with their respective voices on the broader terms of general bodily movement. Because the movement of the character is enough to indicate the source of the voice, *Life is Strange* is able to effectively stitch together its rendered bodies and voices in such a way as to make its conversations easily comprehensible. Even in the modern videogame, fully synchronous lip synching is a virtuosic flair rather than a necessity.

Fig 6 (left) and Fig 7 (right). Max talks to Samuel the janitor in a shot/countershot sequence.

However, this recoding of the symbolic source of the voice is not without consequence. Something of a tension arises, as game voices are imbued with an inherent looseness, despite the
fact that these voices have, for all intents and purposes, been successfully de-acousmatized. Having been granted freedom from the constraining focal point of the mouth, voices tend to hang in the air around characters, contributing to a paradoxical situation wherein videogame voices take on a kind of independence whether or not their source has been identified. This creates a lingering sense of artificiality: voices are attached to their sources, but never in such a way as to let the unified “being” feel completely human. But this lingering tension doesn’t simply manifest as an inherent flaw of the videogame medium: rather, this medium-specific aural independence has given game developers the leeway necessary to develop methods of storytelling that privilege audio, effecting the player’s understanding of time and space in profoundly varied ways. This essay identifies three primary modes of aural-primary storytelling: the Audio-Log/Monologue, Travel Audio, and Potentially Offscreen Bodies. The distinct dissimilarity between these methods renders an overall chronotopal analysis of the aural-primary genre of videogame narrative essentially impossible (a chronotope that encapsulates all of these methods would prove too broad to be useful). Rather, each of these modes will be addressed separately, with issues of time-space referenced only when relevant.

IIa. Audio-Log and Monologue

The cyberpunk action-role playing game System Shock (Looking Glass Studios, 1994) begins with a hacker being captured whilst attempting to gain access to restricted files concerning the Citadel Space Station, which is owned by the TriOptimum Corporate Network. After the hacker’s capture, he is offered a deal by a TriOptimum executive: remove all ethical constraints from an Artificial Intelligence named SHODAN, which controls Citadel Space Station, and all of the charges against the hacker will be dismissed, and he will receive a new neural interface. The hacker does as he is bid, and is placed into a coma to allow for the purpose
of implanting the neural interface. Six months later, the hacker awakes from the coma to find the Citadel Space Station in a state of decrepitude: lights flicker and dead bodies litter the floor. The player is given control of the hacker, and tasked with finding a way to exit the room the hacker has found himself locked inside. To do so, the player must open a supply closet, where he/she will find the key required to exit the room. More importantly, however, the player will find an item called the Multimedia Data Reader, which allows the player to listen to e-mail messages and audio logs. Upon picking up the data reader, the player receives an audio e-mail from Rebecca Lansing, who explains the basic situation on Citadel Space Station to the hacker, and what must be done to stop SHODAN (Holmes 53).

This initial voice-recorded e-mail is the model by which all interactions with System Shock’s non-player characters occur: that is, every interaction plays out as a one-sided conversation wherein the player/hacker listens to an email sent by a survivor on the station, or an audio log found within the environment. There are no conversations in System Shock: even the survivors will be killed before the player has a chance to speak to them in person. The entire narrative is communicated via e-mails – which are either presented linearly based on the player’s progress – or via audio recordings found in the environment, all of which predate the waking of the hacker from his coma and are found non-chronologically (Holmes 54). The game’s producer, Warren Spector, has stated that this narrative design developed out of dissatisfaction with the branching conversation options that were typical of System Shock’s contemporaries which, according to Spector, reduced human conversations to interrogation-like puzzles (Holmes 55). By eliminating traditional dialogue sequences and using audio-logs as the only means of communicating narrative information, Looking Glass Studios developed a narrative method which didn’t interfere with gameplay flow (because the audio logs could be listened to during the
real-time gameplay), and allowed for a non-linear, non-chronological telling of plot events (Holmes 55).

The two kinds of audio recordings the player could acquire in System Shock would diverge somewhat into their own subgenres within this mode of narrative storytelling: the linear, player-progress based emails would develop into what this paper calls “monologue,” while the non-chronological recordings the player can find within the environment would become the foundation for “audio log” narrative. The former would come to be associated with a particular brand of mechanically minimalist art games, typified by works like Dear Esther (thechineseroom, 2012) and The Old City: Leviathan (Post Mod Softworks, 2014). In First Person Walkers like these, the player walks through a setting, triggering spoken narration as he/she proceeds through certain points in the game environment. Typically, games like Dear Esther and The Old City have few (or no) NPCs for the player to encounter, and only one narrative voice. The “audio-log” mode of narrative, on the other hand, is now primarily associated with the work of American auteur Ken Levine, who served as the lead writer and designer of the direct sequel to System Shock, as well as the game’s spiritual successor, Bioshock (2K Boston, 2007). Levine’s work would follow the basic structure of System Shock very closely, often mixing in-perspective, linearly presented cinematics and one-sided radio conversations in conjunction with optional audio-logs.

However, despite this slight divergence, the core of the monologue and the audio-log narrative remains the same: these modes are marked by extensive use of purely acousmatic sound. Because neither game ever deviates from the first-person perspective, the source of the narrative voice of Dear Esther and The Old City is never revealed, and games like Bioshock rely on diegetic radios or similar non-visual media players, which are inherently acousmatic.
instruments seeing as they “transmit sounds without showing their emitter” (Chion, *Audio-Vison* 71). This heavy usage of voices without sources generates a complicated relationship between time, space, and narrative, insofar as the narrative information has very little impact on the construction of a space, and in the case of audio-log narratives, has at best an indirect relationship with the player’s experience of in-game time.

There is, however, a sense in which this essay’s notion of player-action time can be applied to the monologue (and to a lesser extent, the audio-log). Because the monologue narrative is constructed as a series of voice over fragments, with each fragment triggered by reaching a certain area within a largely static game environment, the basic model of player input → plot output that serves as the foundation of player-action time remains more or less intact. However, the way in which monologue narratives do not interrupt the basic flow of gameplay complicates this basic construction of time: there is no segmentation of time, wherein looped-time is interrupted and moved forward for the sake of plot progression. Rather, monologue fragments are presented concurrently with the player’s movement through the world, unaccompanied by any kind of visual plot event that would serve to interrupt that flow of play by either diverting attention or removing player agency for an extended period of time. This separation between the aural fragments and the pacing of the player’s movement through the environment essentially surrenders the pace of the narrative wholly to the player, as there is no other diegetic entity in a basic monologue narrative to contest the player’s experience of time.

The resulting experience of time is typical of what Mark Wolf identifies in games with fluid movement through a space: the close correlation of game time and real time, wherein time experienced by the player and the player-character is fairly congruous (86). Audio-log narratives work in a similar manner, save for the addition of one clearly defined plot event: interacting with
the recording itself. This simple interaction creates an experience more akin to the player input → plot output foundation of player-action time. However, the actual experience of the the audio-log happens during a moment of unbroken player agency and undirected attention, just as in the monologue. Ergo, what the player experiences between the recoveries of each separate audio log isn’t looped time, as these periods of time are filled with narrative information upon the retrieval of an audio-log. The player’s experience of time in the audio-log narrative is the same as the monologue narrative: a congruous experience of time between player and player-character.

The relationship between the monologue/audio-log mode of aural narrative and the player experience of videogame space is even more ill-defined. Because the voices of the monologue and audio-log remain firmly attached to the player, the aural information takes no part in the construction of the visual space the player is navigating. Thus, the relationship between this mode of narrative and the space it occupies is purely contextual, wherein the narrative fragments the player hears while traversing a space provide some kind of backstory or thematic framework for said space. For example, a winding staircase in The Old City: Leviathan is linked, via monologue, to the relationship between evil and idiocy, and in Bioshock, two corpses lying on a bed become, thanks to an audio-log left at their side, two bereaved parents who failed to cope with the loss of their daughter. This mode of aural narrative essentially offers additional dramatic context for the space in which the player finds themselves. Monologues and Audio-logs do not construct spaces, but enrich them.

IIb. Travel Audio

In the opening cutscene of Grand Theft Auto IV (Rockstar, 2008), protagonist and player-character Niko Bellic steps off the boat into Liberty City\textsuperscript{5} for the first time with little more than a suitcase in hand, awaiting the arrival of his cousin Roman, whom he’ll be living with. Niko

\textsuperscript{5} The Grand Theft Auto series’ fictitious facsimile of New York City
spends a few moments alone in a dreary industrial dock, the lights of the city obscured by fog and massive tankers quietly floating in the port. Soon enough, the sound of distant Russian pop music pierces the air and a lone car can be seen careening down the dock towards Niko. Upon reaching Niko, the humble cab screeches to a halt, and out of the driver’s seat a drunken Roman emerges, yelling ecstatically about the arrival of his beloved cousin. The two men share a fond hello and, as Roman stumbles about, Niko offers to drive the pair of them home. The two men enter the car, and the cutscene ends: the game cuts to a long shot behind the car, and the player is given control. As the player presses a button to make the car sputter to life, Roman tells Niko where to go, and a route is outlined on the extradiegetic mini-map in the corner of the screen. As the two men drive off towards Roman’s apartment, they begin to reminisce about the past, and express excitement about their future together in America.

This short sequence serves as a prototypical example of the structure of Travel Audio: a conversation between at least two characters, whose voices have both been previously decousmatized, as they travel between two points in a playable sequence. This kind of sequence is synonymous with Rockstar Games’ open-world titles, which effectively utilize the extensive travel time inherent in the genre in a way most open-world titles fail to do, using these slow moments of transit as an opportunity to further the narrative. Travel audio is distinguished from the other modes of aural-primary narrative identified in this chapter by the way in which characters are blocked in these moments of travel: while the characters engaged in these travel audio sequences are not necessarily onscreen – as their figures are often obscured by the body of the automobile they are sitting in – the player is still aware that all of the source of every voice they can hear is situated within the frame. Because the characters are blocked closely together, this aural mode creates a singular experience of videogame space.
Michel Chion posits that the spatial relationship between the voice and image is largely uncertain: that “in fact the greatest arbitrariness does prevail with regard to space” (*Voice In Cinema* 129). He suggests that spatial incoherence does not bother the audience, specifically citing the way in which film has normalized presenting characters in long shot while playing their voices as though they were close to the camera (*Voice in Cinema* 129). Interestingly, it is this exact arrangement of voice and image that defines Travel Audio: characters gather into a vehicle, which is then shown in long shot (by default) during gameplay, while the dialogue is mixed in such a way as to dominate the aural environment making it appear as though the characters are, to borrow a film phrase, miked closely. It is worth calling into question the arbitrariness of this relationship between image and voice, as the nature of the videogame medium raises questions of player identification that may be resolved by examining the closeness of the voice in Travel Audio.

Frank Biocca has posited that when a user interfaces with a virtual environment, his/her sense of presence actually oscillates between three different environments: the physical environment (distal immediate), the virtual environment (distal mediated) and the imaginal environment (minimal attention to distal stimuli). Furthermore, within this virtual environment, there are three bodies present: the objective body [“the physical, observable, and measurable body of the user”], the virtual body [“the representation of the user's body inside the virtual environment”], and the body schema [“the user's mental or internal representation of his or her body”] (Biocca, Cyborg’s Dilemma). Lori Landay has adapted this assessment and reconfigured it for player engagement in videogames, wherein the three bodies present are the *physical body*,
the *screen body*, and the *imagined body*\(^6\) (19). Landay suggests that sound has the potential to bridge the gap between these levels of presence in a way that image and action cannot: in an examination of Limbo, Landay identifies how “although we are always visually distant from the [avatar], the sound is immediate” (18). It is the immediacy of this sound (as well as, in the case of Limbo, that sound’s musical quality), Landay suggests, that “connects the player to the physical and emotional experience of the avatar body” (19-20). The closeness of the voices of Travel Audio captures that same sense of immediacy, re-positioning the player within the onscreen space. The close proximity of the voices despite the use of the long shot serves to bridge the gap between the physical body and the screen body, creating a dualism in the imagined body wherein the player is simultaneously positioned within the player-character’s means of transit while still being physically removed from it via the third person perspective. Thus, Travel Audio is able to reconfigure videogame space in such a way as to let the player align him/herself with the screen body, relocating him/herself within an environment and transforming the sense of control over an avatar to a sense of presence within said avatar.

The simplest way to understand the way in which Travel Audio effects the progression of time in videogame narratives is to actually remove the Travel Audio from its own schema. Without the audio component, what remains of the basic narrative structure outlined previously is an uncomplicated example of player-action time: two moments of plot progression (the respective events that occur at points A and B) are separated by bouts of cyclical time (this time is often marked by an inconsequential day-night cycle in open world videogames), wherein the player is able to further the plot by traveling to point B. Travel Audio complicates this basic adherence to player-action time, as it fills the cyclical time with consequential character

\(^6\) In my understanding, this terminology deemphasizes the way in which the virtual body, in Boccia’s usage, can interfere with the sensory feedback of the objective body, and is thus more appropriate for the simpler interactions players have with the virtual environments of videogames
interaction. In these instances, then, the player’s experience of time is comparable to the player-character’s experience of time, meaning that travel audio injects the congruous time of the monologue/audio-log into the basic formula of player-action time. The aural construction of congruous time overrides any visual cues of cyclical time, thus privileging the audio, and reconfiguring the player’s understanding of the progression of in-game time as something approximating the time experienced by the physical body.

IIc. Potentially Offscreen Bodies

Referring to this mode of aural narrative as “Potentially Offscreen Bodies” is something of a simplification (this designation could easily refer to monologue/audio-log narratives, given that mode’s use of acousmatic voice), though it does encapsulate this method’s central conceit. The name “Potentially Offscreen Bodies” refers to the possibility that, during a gameplay sequence in which the player is given full control of the camera, the player will frame their own perspective in such a way as to remove a non-player character is who actively engaged in some form of conversation, either with the player or another NPC, from their field of view. In such a scenario these characters, who are necessarily located within the immediate area, are relegated to the videogame’s “offscreen” space. This mode of aural mode occurs only in sequences wherein the player is given full control of the in-game perspective while the de-acousmatized voices of characters visible within the immediate area are audible: these occurrences may or may not consist of scripted sequences of movement on the part of the non-player characters, but in either case, the player’s perspective is not forcefully directed at these nearby characters. A prototypical example of this type of sequence occurs in the early hours of Half-Life 2 (Valve 2004), in which the player is able to freely walk around and explore the lab of one Dr. Kleiner, as the doctor and the player’s companion, Alyx Vance, discuss a teleportation machine that the pair plans on using
to safely travel across a dystopic, 1984-esque cityscape. This conversation features scripted movement of Vance and Kleiner and directly informs what is to come in the game’s narrative, but the player is free to focus on the details of their choosing in the lab. This surrendering of the visual focus to the whim of the player implicitly prioritizes the authored sound over any authored image (as the sound of these authored sequences is designed to be heard whether or not the attention of the player has been diverted), meaning that this kind of narrative sequence is aural primary.

It is this variation of aural narrative that bears the closest relationship to the player’s experience of videogame space. Mark Grimshaw has noted the way in which sound operates in both the physical and virtual environment, because sound “has a real volume and dimensionality that is a 3-dimensional representation of the 2-dimensional representation of the 3-dimensional world of the game” (119). Because of the 3-dimensional nature of sound, says Grimshaw, videogames “posit the player as a first-person auditor,” insofar as the diegetic sonic environment of the videogame “extends from the screen to physically encapsulate the player in the acoustic ecology’s real resonating space” (122). Grimshaw makes this argument specifically in the context of First Person Shooter games, but his observations are applicable to genres beyond those that literally situate the player in a first-person perspective. The way in which Travel Audio alters the player’s understanding of aural space in order to let him/her identify with the screen body provides us with a useful example of the way in which players act as first-person auditors even when the game is presented via a third-person visual perspective: if the player hears the audio from the relative perspective of the screen body, then the player still takes on the role of first-person auditor. It is this role, combined with the use of de-acousmatized voice, that allows the player to experience the three dimensional sound-space in a way that is disallowed by the
aforementioned modes of aural narrative. Because the player is allowed to freely move their character within the visual environment, he/she gains a measure of kinaesthetic control over the game’s aural playback: the volume/direction of the sound changes as the player-character’s position moves in relation to the source(s) of the audible voice(s) (Grimshaw 120). This capability allows the player to use the narrative audio as a means to discern their position relative to any other characters occupying the same space. This use of offscreen de-acousmatized conversation allows the player to place themselves within the same relative position within a virtual environment as the screen body they control, contributing to the player’s identification with said body, allowing for more effective interpellation of the player when in-game characters speak directly to the player-character he/she is controlling.

The experience of time in this narrative method is most akin to that of the monologue/audio-log, as the unbroken flow of gameplay disallows the kind of temporal segmentation/differentiation that serves as the foundation of player-action time. Just as in the monologue/audio-log method of narrative, the use of speech confirms the player that they are experiencing a linear progression of time. Furthermore, the way in which the player is cast as a first-person auditor during scenes of Potential Offscreen Bodies acts as proof that that he/she is embodying their screen body, meaning that the player’s experience of aural stimuli mirrors the hypothetical experience of the player-character. Ergo, the linear progression of time the player experiences is the same as the in-game time. This final aural mode of narrative is marked, just as the previous two, by the use of congruous time.
III. Image-Primary

It would be little more than a slight exaggeration to posit that videogame academics are profoundly terrified by images. Videogame studies has been largely defined by a rejection of the image, based primarily in the early conflict between the “‘ergodic’ and the narrative” as precipitated by the cotemporaneous publishing of Espen Aarseth’s *Cybertext* and Janet Murray’s *Hamlet on the Holodeck* (Cheng 15). This conflict would serve as the basis for arguments that videogames cannot be considered one unified medium based on a supposed lack of a uniform aesthetic (Aarseth Year One); or that because of the various representational strategies at work in videogames, any discussion of them as a medium must deploy interactivity as their singular commonality (Apperly 7). A similar line of logic would be utilized in James Newman’s infamous argument that the over-sexed representation of *Tomb Raider*’s Lara Croft does not matter because the “the pleasures of videogame play are not principally visual, but rather are kinaesthetic.” Aarseth himself would argue that the dimensions of Lara Croft’s body are irrelevant because they do not affect the way the act of play: “I don’t even see her body,” says Aarseth, “but see through it and past it” (Genre Trouble). Somewhat ironically, these kinds of arguments can be understood as being rooted in a pictorial turn in contemporary criticism, wherein images “are all-powerful forces… or they are denounced as mere ‘nothings,’ worthless, empty, and vain” (Mitchell 77). One can plainly see the real fears of radical ludologists in Mitchell’s jocular summation of one of the underlying motivations of this pictorial turn: “[images] seem to simulate everything, and therefore they must be exposed as mere nothings” (76). The irony of Aarseth arguing that the computer game is a genre of simulation while dismissing the importance of the image is not lost. Of course, this impulse to dismiss images is not universal among videogame academics: Mark JP Wolf posits that videogames “are best
approached and analyzed using tools developed in film and television theory and media studies” (2). It is with this precedent that this essay consciously rejects the aforementioned radical ludologist attitude that has been largely canonized in game studies, in favor of an acceptance of existing theories of visual media.

All this is to say: the image matters. Still, it’s necessary to elucidate what this essay is referring to as the “image.” The use of the word image in the context of this essay is, admittedly, rather reductionist: images are “immaterial symbolic forms” that can correspond to many forms of material art, which is why we can speak of architectural, sculptural, cinematic, textual, and mental images (Mitchell 85). One could easily argue that a text adventure game like *Planetfall* is constructed of multiple sorts of images: the mental image that emerges in the player via the textual descriptions of the diegetic world, and the in-game text as an image onto itself: the color and font of which constituting their own identifiable aesthetic. However, this usage of the term is beyond the scope of this chapter, which is interested in images in videogames as a mode of pictorial representation: that is, this chapter is interested in the way in which objects are rendered, positioned, and framed within a videogame in order to function as a narrative text. To this end, this chapter will focus on the way in which videogames have adapted mise-en-scène and pre-determined framing as a means of communicating narrative information at the level of pure image, or via the remediation of film. The image-primary narrative genre, therefore, is deployed in two primary modes: that of the embedded narrative, and the cutscene.

### IIIa. Embedded Narrative

The term “embedded narrative” was coined by Henry Jenkins as a part of his article “Game Design as Narrative Architecture.” Jenkins’ notion of embedded narrative is founded,

---

7 This accusation is largely based on this author’s anecdotal experiences. However, Google’s citation statistics for Aarseth’s *Cybertext* and Wolf’s *Medium of the Video Game* indicate that the former has seen far more proliferation than the latter.
first and foremost, in the basic distinction between plot and story, wherein plot is the structured set of causal events as they are presented in a work, and story refers to the audience’s mental construction of the chronology of those plot events (Jenkins 126). Based on this model, “narrative comprehension is an active process by which viewers assemble and make hypotheses about likely narrative developments on the basis of information drawn from textual cues and clues,” in which the audience is engaging with a story that is “less a temporal structure than a body of information” (Jenkins 126). In a videogame, argues Jenkins, this non-temporal story information can be distributed across the game space, creating a narrative that is “prestructured but embedded within the mise-en-scène awaiting discovery” (126).

The first person exploration/puzzle videogame *Ether One* (White Paper Games, 2014), provides a useful demonstration of Jenkins’ embedded narrative, as the game is heavily invested in the player engaging with and extrapolating information from details within the mise-en-scène. In *Ether One*, the player takes on the role of a “restorer,” a person who has been hired to participate in an experimental project (called the Ether Project), which involves entering the memories of an old woman named Jean, who is suffering from dementia with the intent of piecing together the old woman’s memories and curing said illness. The woman’s memories take the form of a simulated version of her childhood home, a small British port town called Pinwheel, where the restorer is tasked with collecting memory fragments (which take the visual form of ribbons) and interacting with the environment in such a way as to re-enact the daily lives of Pinwheel’s inhabitants, which leads to the reconstruction of in-game projectors which give the player more information about Jean’s treatment (but ostensibly has no effect on the restorer’s central goal of curing Jean’s dementia).
Structurally, *Ether One* resembles the detective stories that Jenkins cites whilst demonstrating the normality of semi-nonlinear narratives. Detective stories, explains Jenkins, tell two stories: one chronological (the investigation) and one told out of sequence (the events of and leading to the crime that is being investigated) (126). *Ether One*’s narrative is based around a similar duality, telling a more or less linear story of the Ether project and the project lead’s relationship with Jean and the Restorer while simultaneously telling the story of events leading up to and following the collapse of Pinwheel’s ore mine, and that event’s subsequent, devastating impact on the insular community. The former is an aural-primary narrative, told in the mode of the monologue as the acousmatic voice of the Ether project’s director speaks to the Restorer as they recover memory fragments; the latter largely takes the form of embedded narrative, with the most intimate details of the general population’s lives hinted at within the mise-en-scène.

An early example: as the player walks through Pinwheel Harbor, they may happen across a two-story residence with a particular prevalence of medical equipment, including IV bags and rolling walkers. Should the player explore the rest of the bottom floor, he/she would find a bedroom with a single bed, a television set, and more medical equipment (fig 1), and a table set for two (fig 2). From this information, the player can ascertain that whomever lived in this home was elderly and in need of consistent medical treatment, which appeared to be administered by someone else: the question of who exactly that person was remains unresolved; until, that is, the player explores upstairs, where he/she will find another bedroom, where he/she will find one fully furnished bed accompanied by one empty bedframe (fig 3). From this new information, a concrete chain of events becomes evident, wherein the layout of this home has been adjusted to accommodate the illness/age induced physical disabilities of a spouse. The specifics of the relationship between the afflicted and his/her caregiver are unresolved, but one can easily picture
the quietly tragic sight of one spouse assisting the other to the dining table for tea, a formerly idyllic scene warped by the addition of medical equipment, and the knowledge that the two will sleep, like they have every night since the bed was moved, in separate rooms.

![Fig 1: The living room](image1)

However, *Ether One* has use as a prototypical example of embedded narrative beyond its attempts at emotional affect: rather, it is the game’s premise and narrative structure that allow it to deploy embedded narrative in such a way as to make this storytelling mode’s effect on the videogame time-space explicit. The town of Pinwheel in *Ether One* is a temporal and spatial oddity. It is a uniquely false environment because it is *diegetically* false, a simulation generated by an unreliable mind: the way in which Jean’s nostalgia filters the visual representation of her hometown is immediately evident in the game’s painterly style, which gives even the dank Devlin Mine an aesthetically appealing sleekness. More importantly, though, the notion that Pinwheel as the player experiences it is not a “real” environment, but an amalgamation based in
decades of experiences of a single person, means that the environment is temporally condensed: while the player experiences Pinwheel as a series of continuous spaces, the various parts of those spaces do not exist within the same temporal moment. The town of Pinwheel, as the player experiences it, is a place where diary entries written by Alexander Graham Bell as he first conceived of the telephone exist concurrently with notes written between two aging lovers in the latter half of the 20th century. The synchrony of these two kinds of objects obfuscates the player’s own experience of time in relation to the world around them: any tangible linkage between the player’s experience of time and the passing of time within the diegetic world appears to be severed. Pinwheel, as it is represented in *Ether One*, lacks a temporal present: it is a reflection, a collage of still lifes that creates the impression of a place that has existed. This is an experience at odds with most videogames, which seek to create worlds that have a ‘present moment’ in which the player-character exists and a linear chronology. Pinwheel lacks these traits. It would not be inaccurate to call Pinwheel a dead environment: it is a collective of frozen objects, all trapped in their own unique moment in time. It *would* be inaccurate, however, to say that time in Pinwheel is achronic: that is, the embedded narratives of *Ether One* are not “deprived of every temporal connection” in such a way as to be “dateless and ageless” (Genette 84). These vignettes historicize Pinwheel, creating a place with a tangible past, albeit a past that is expressed through static objects that are unaffected by the passing time that the player experiences: As such, *Ether One* can be best described as being polychronic, a form of time in which “the narration anchors events in multiple temporal frameworks and thereby promotes competing ways of sequencing those events” (Herman Polychronic 75). As the player works their way through the environment of *Ether One*, competing linearities of narrative arise: are the

---

8 Jesper Juul makes a profoundly problematic argument that because interactivity and narration cannot co-exist, games are almost always chronological (*Games Telling Stories*). The author would like to acknowledge the similarities of his and Juul’s claim, while distancing himself from Juul’s overall argument.
vignettes embedded in the mise-en-scène ordered based on the way the player discovered them, or based on their actual position within the historical timeline of Pinwheel itself? Due to the lack of a true “present” in *Ether One*, this question remains unresolved, and the order of the vignettes can be best defined as being “partially ordered”: that is, the events of *Ether One* can be “uniquely sequenced relative to all others, some only relative to some others, and some relative to none” (Herman, *Story Logic*). The embedded narratives that serve as steps in puzzles are relative to the other steps in those same puzzles, while other embedded narratives exist entirely independent of one another, their relative positioning in the timeline of Pinwheel remaining largely undefined.

While Pinwheel itself lacks any kind of tenable temporal present, *Ether One* has to possess some kind of “present moment” in which the player can navigate the game world: because the player can actively control a character in the game environment, there is a sense in which the story time is synchronous with the “reading/viewing time” (Juul, *Games Telling Stories*). However, because Pinwheel lacks any kind of diegetic present, this supposedly synchronous time is frustrated: the game world does not progress temporally at a rate analogous to the reading time the player experiences. Through repetitious use of musical pieces and sound effects, as well as consistent environmental animations, *Ether One* is able to create the impression of a place that appears active whilst the player “reads” it, but is functionally frozen. Therefore, the “present” that the player occupies while playing *Ether One* operates on the logic of Mark Wolf’s looped time: that is, again, a combination of stillness and movement in which ambient sound and motion emphasize the potential for player action and keep a scene feeling “live” (80). The looped time in *Ether One*, with the exception of a few scripted story sequences, remains unbroken because the premise demands that the world not progress with the actions of
the player: every narrative embedded within the mise-en-scène has already occurred, giving the player no agency in regard to its outcome.

It was mentioned previously that *Ether One* is not unique in the nuances of its deployment of embedded narrative, but simply in how it brings the basic mechanics of the embedded narrative to the forefront of its presentation. Simply put: embedded narratives are historicized, polychronic vignettes that are contained within segments of looped time. Because they exist solely within the mise-en-scène of a game environment, embedded narratives must take place prior to their discovery by the player. They lie in wait within an environment, like a fallen tree in a forest: the tree fell before our discovery of it, but we are left to imagine the sound it made. But because the player misses the actual event that the scenery alludes to, there is no biographical plot event to break the sequence of looped time. Embedded narratives, one could say, are not simply “embedded” into the mise-en-scène of the game environment, but are also “embedded” into the act of navigation. That is, embedded narratives are situated within the environment in such a way as to not interrupt the basic flow of gameplay, insofar as they are designed to be looked at and not actively interacted with, meaning that these embedded stories are contained within looped time sequences of play, rather than acting as interrupting moments of biological significance to the playable character (though occasionally, a character seeing part of an environment that corresponds to an embedded narrative will trigger a segment of dialogue – which would fall under either the monologue or Potential Offscreen Bodies modes of aural narrative). Additionally, embedded narratives are polychronic insofar as they are partially ordered, especially when considered relative to one another: embedded narratives are always ordered as having happened before the current moment in which they are discovered, but are often indeterminately related to one another. This essay’s prior referral to Pinwheel as a “dead”
environment was not an unconsidered choice of words, as many embedded narratives are quite literally dead. Take, for example, a darkly humorous scenario found in *Fallout 3* (Bethesda 2008), in which a crashed motorcycle can be found in a large tunnel, next to a car, close to which wooden boards have propped up at an angle, as a makeshift ramp. Above the car, half of a human skeleton is embedded into a hanging light (fig 4). The events that led to this arrangement of objects within the gameworld should be obvious. But the events themselves are only partially ordered: the pre-figuration of the environment (as well as the skeleton) obviously codes the events as having happened prior to the arrival of the player, but the level of decay renders any other conclusions about the timing of the motorcyclist’s grisly end relative to other past events unknowable: ergo, the scene is partially ordered, placed on a timeline relative to the player’s narrative, but not to other elements of the gameworld’s past, rendering the scene polychronic.

![Fig. 4: The motorcycle crash. Credit for screenshot goes to reddit user “PrincessHorse” who posted the above photo to an online forum, with the title “Fallout 3 has many silent stories being told. Just stop and look around”](image)

The embedded narrative’s relationship to space is, in certain ways, similar to its relationship with time. Most importantly, as has already been indirectly stated above, embedded narratives historicize spaces, imbuing them with a backstory that makes navigation more pleasurable, insofar as it imbues play with thematic and narrative significance it would lack otherwise. It does so by essentially performing the reverse of what Wei, Bizzocchi, and Calvert
describe as the function of characters attached to one space: characters that are spatially at rest rather than in movement. These kinds of characters, Wei et al. argue, become the background of a space, or part of that space’s context, especially when not interacting with the player: therefore, “these characters play the same role as other environmental objects” (10). Logically, embedded narratives, which are expressed via environmental objects, can thus be said to play the same role as characters – especially considering the way in which embedded narratives are often the residual effect of the actions of a character that occupied the space at some point in the past. Embedded narratives, just like background characters, become a part of the space’s context.

More specifically, embedded narratives influence a space’s immediate context: Wei et al. note the way in which interactions with “background” characters, the plot can change locally – in opposition to mobile characters, whom can affect greater change in a videogame’s plot (10). Just like these characters, embedded narratives are spatially at rest, affecting the “plot” in the most immediate sense, at the level of a vignette that is tied to a single area.

IIIb. The Cutscene

In Remediation: Understanding New Media, Jay David Bolter and Richard Grusin outline the “logic of remediation,” which is “the formal logic by which new media refashion prior media forms” (273). Bolter and Grusin identify two primary strategies of remediation, which Thomas Apperly effective summarizes (in the midst of an argument against them):

“Bolter and Grusin (1999) describe two strategies of remediation: transparent immediacy and hypermediacy. Transparent immediacy seeks “to get to the real by bravely denying the fact of mediation” (p. 53). Hypermediacy—conversely—draws attention to the act of mediation, “by multiplying mediation to create a feeling of fullness, a satiety of experience, which can be taken as reality” (Bolter & Grusin, 1999, p. 53)” (7).

The cutscene can perhaps be considered the epitome of hypermediacy in the context of the videogame. It is, for all intents and purposes, pure imitation: a transplantation (of varying elegance) of the rules and conventions of a whole other medium into the language of the
videogame. The cutscene, a common term within videogame parlance, is effectively defined by Dylan Holmes as a “scripted scene of exposition that the player can’t control, resembling a short film” (211). It is this lack of control that has made the cutscene such a reviled entity in modern videogame discourse: Jesper Juul posits that “cut-scenes are often considered problematic because they prevent the player from doing anything and are in a sense a non-game element in a game” (qtd. in Cheng 16). Torbin Grodal characterizes the cutscene – or “film-sequences,” as he calls them – as actually blocking interaction, making the cutscene a more “dead” experience than its filmic counterpart (144). However, since this essay accepts Wolf’s position that games can be safely approached via film theory, this essay accepts the logic of remediation as a useful means of understanding the role of cutscenes in the medium of the videogame.

Hypermediacy is a useful concept in understanding the way in which cutscenes function in the modern videogame, insofar as it explains the lengths videogames go to imitate the basic language of cinema. Take, for example, the opening cutscene of Metal Gear Solid 4 (Kojima Productions, 2008) which, from its very first moments, makes its investment in cinema explicit. The very first in-game shot (MGS4 actually begins with an extended Full Motion Video sequence, featuring real actors in fictional television shows that exist within the game’s diegetic world) is a shot of the sun (fig 5), which serves not only to establish one of the game’s recurrent visual motifs, but also showcases the extent to which the developer has gone to make the cutscenes feel “cinematic.” The developer has attempted to capture that sensation of cinema by imitating the imperfections and nuances of the film medium’s most vital piece of technology: the camera. Pointing the “camera” directly at the sun allows MGS4 to showcase several of its filmic visual effects: immediately apparent is the addition of lens flare, as well as the illumination of several “scratches” on the “lens.” As the camera tilts downwards it jitters and shakes
imprecisely, imitating a handheld cinematographic style. In the next shot (fig 6), the player can see the way in which \textit{MGS4}'s camera imitates the focal length characteristic of a normal lens by having the foreground of the frame remain in focus while the background is out of focus. The basic goal of hypermediacy, creating a “full” experience that can be taken as reality, immediately elucidates the purpose of these visual effects. The imprecise camera movements and imperfection of the image are, in film, coded as “real”: these are the same visual cues that lend a film like \textit{Children of Men} (Cuarón, 2006) their intense, documentary-esque aesthetic. All of these “realistic” visual cues are reliant on the presence of a camera, a technology which games inherently lack. In deliberately simulating the aesthetic of the film camera, \textit{Metal Gear Solid 4} is enacting the strategy of hypermediacy, insofar as it is drawing attention to the medium that is being refashioned by consciously imitating the visual indicators of the artifice of the camera. In essence, because the modern cutscene is a remediation of film, they behave as though there is a real camera present, “filming” the action the player watches. The rules (e.g. continuity editing) and aesthetics of film carry over to the videogame: film realism is videogame realism.

One particular element of film gets reconfigured significantly in the context of the videogame cutscene: the edit. In an article on the interactive drama/thriller \textit{Heavy Rain} (Quantic Dream, 2010), Ian Bogost posits that the central difference between the film and videogame medium is that the “primary quality” of the former is editing. Bogost’s claim about the essence
of cinema lying in the cut is well-founded: he argues that Sergei Eisenstein believed that editing techniques “made it uniquely possible for cinema to link unrelated images through juxtaposition,” and cites Francis Ford Coppola’s testimonial that “the essence of cinema is editing” (“Francis Ford Coppola Interview”). Bogost then argues that videogames lack cinematic editing “because continuity of action is essential to interactive media,” and from this observation, he posits that the essence of gameness may lie in the “rejection of editing in favor of prolonging.” For evidence, Bogost turns to two gameplay sequences from Heavy Rain – a search for a child in a crowd and a quiet hunt for another child’s teddy bear at home – in which the lack of a cinematic edit works to generate a feeling of panic and silent unease respectively. While Bogost’s assessment of these two gameplay sequences is more or less accurate, his argument that gameness lies in prolonging is inaccurate in two ways, the first being that games released after the publishing of Bogost’s article, like Thirty Flights of Loving (Blendo Games, 2012) and The Stanley Parable Demo (Galactic Café, 2013), would break the continuity of action inherent to interactive media by incorporating jump-cuts into interactive sequences acting as an effective counter-argument to the long-held belief that gameplay must be continuous.

Secondly, Bogost’s argument fails to account for the existence of cutscenes, which, as an interruption of gameplay, reconfigure the cinematic cut as something broader. The insertion of a cutscene before or after a gameplay sequence is analogous to the filmic edit: it is an interruption of what would be an otherwise unbroken sequence and its effect is to provide context for the sequence with which it is juxtaposed. This context through juxtaposition is known in film as the Kuleshov effect, named for Soviet filmmaker Lev Kuleshov, who “demonstrated that the manipulation of context can alter an audiences’ perception of an actor’s facial expressions, thoughts and feelings.” Kuleshov juxtaposed identical shots of Soviet actor Ivan Mozzhukhin’s
face with a shot of a funeral, or a happy child, leading people to conclude that the actor was either expressing melancholy or happiness, respectively (Mobbs et al). The relationship between a cutscene and the gameplay sequence it is juxtaposed with is analogous to the relationship between the shot of the happy child and the face of Ivan Mozzhukhin: the latter acts as context for the former. Cutscenes are not merely “dead” interruptions of play – they are very much alive, informing the play that surrounds them, doling out information on the motivations of the player-character (and thus, to the player him/herself). *Metal Gear Solid 4* actually makes this connection explicit by manipulating the in-game meters that represent the protagonist’s physical and mental health based on what the protagonist experiences during cutscenes. In doing so, *MGS4* explicitly demonstrates its continuity between cutscene and gameplay. This continuity is always present in any videogame, though, because the cutscene remediates the very essence of film, the edit, by juxtaposing two modes of engagement (rather than two shots) in order to make one give context and meaning to the other.

It’s the particular mode of engagement that the cutscene represents that informs the way in which the player experiences time. It is difficult to ascribe one single experience of time to the cutscene, seeing as the form adopts complex rules of cinema, a medium which has the capability to represent time in myriad ways. However, by viewing the cutscene as a part of the larger structure of a videogame, we can examine the medium-specific way in which this mode of narrative influences the player’s perception of time. Essentially, the cutscene is the antithesis of the embedded narrative; insofar the two modes occur within opposite moments of player-action time: the latter is contained within the looped time of gameplay, whereas the former is the interruption of that looped time. The presence of a cutscene, as the prototypical “non-interactive” narrative device of the videogame medium, always signals a moment of biological importance,
relevant to the progression of the videogame’s plot. Therefore, the experience of time in a cutscene is more likely than not an experience of forward temporal motion, especially if the cutscene in question is invested in the rules of Hollywood continuity editing (and given the remedial-at-best cinematography at work in most videogames, this is likely the case). The use of continuity editing (i.e. following the 180 degree rule, constructing dialogue in terms of shot/reverse shot sequences, etc.), especially in the case of a cutscene that follows a conversation, often suggests a synchronous progression of game time and real time. However, in-game cinematics never strictly generate a sense of synchronous time. The use of filmic conventions allows for time-condensing ellipses, for example, but even more problematic is the inclusion of interactivity in what appears to be an ordinary cutscene: as talked about in the first chapter of this essay, giving the player agency over in-game dialogue injects the logic of player-action time into the cutscene itself, as the game awaits the player’s input in sequences of looped time. Interestingly, the inclusion of agency into a cutscene via a “quick time event” (in which a contextual button prompt appears on screen, indicating that the player must hit said button or be confronted by some kind of failure state) often reinforces the sensation of synchronous time. The implicit time pressure of the quick time event conflates the time experienced by the player and the player-character, insofar as the two have the same amount of time to react to whatever situation they are in. But this experience of synchronous time is not consistent across all cutscenes. What is consistent is the way in which these sequences mark temporal forward momentum, interrupting the looped time of player-action time.

Given the established complexities of film, it is equally difficult to pin down the precise way in which the cutscene effects the player’s perception of videogame space. However, this

---

9 See page TBD in Chapter I
10 There are exceptions to this scenario, as QTEs will often be paired with sequences presented in slow motion
essay has already examined the way in which the hypermediacy of the cutscenes codes videogame spaces as “real”: a particularly interesting strategy given the falseness inherent in the videogame medium. Unlike in the cinema, there is no pro-filmic referent for what the player sees: that is, none of the environments depicted in the game world needed to exist in the “real world” prior to the creation of the game itself (Wolf 51). Wolf notes the way in which this inherent lack of structure means that “space can be shaped or structured in new ways that did not develop in film or television” (52). Wei et al make a similar observation regarding the “virtual camera” of the videogame, noting that the view of a videogame is “computational and dynamic” insofar as it is not “constrained by the laws of optics and physics”: this allows the virtual camera of the videogame to reach “far beyond the functionality of a real camera” (11). Given this knowledge of videogame space, and the tools that games possess to explore those spaces, perhaps cutscenes can be best characterized as imitations of constraint. By imitating cinematic devices and aesthetics, cutscenes are presented with false constraints (like the presence of a camera man, for example) that suggest to the player that the environment he/she is navigating was conventionally filmed. This lends the environments a subtle verisimilitude: because the game space is presented like it was filmed conventionally, the player is given the impression of there being a pro-filmic (pro-ludic?) referent, without there actually being one. Cutscenes, with their hypermediation of cinema, treat digital environments as though they were tangible physical environments, as though the environments did, in fact, predate the videogames themselves.
Conclusion

It is necessary for the sake of clarity for this essay to treat each of the narrative genres it identifies – and the various modes within each of those genres – as though they are inherently designed to function independently from one another. It is certainly true that each of these genres can function as the sole means of transmitting narrative information to the player, the truth of the matter is that two or more of the narrative genres modes described in this essay will, more often than not, be used dynamically within a single work. Such an expectation is built into the definitions of the genres themselves: all three modes of aural-primary narrative are based on the visual framing of audio sources, and the spatiality of the Text-Primary genre’s textual labyrinth can easily be transferred into image-based videogames, implanted into the implicit spatial structure of the branching dialogue tree rather than the spatial makeup of the gameworld. Again, such is the reason for the use of the word “primary” in the designation of these genres: to suggest that there is not just one single operation at work when one of these genres can be identified, but that one might be momentarily privileged over another. Different narrative genres can and will be deployed in different games for the sake of creating a varied experience, fluidly changing the player’s understanding of their position within the game’s time-space as the needs of the work demand it.

Take, for example, the typical structure of most missions from any of the later, three-dimensional Grand Theft Auto games. The player-character will be in control of the player, who is either walking or driving around the gameworld. The player-character receives a phone call from one of the game’s many NPCs, in which that character tells the player’s avatar that there they have some kind of assignment for them to do. After this exchange, an icon appears on the in-game map, informing the player that a new mission is available, and will begin when the
player arrives at the indicated location, where the NPC who just called is likely waiting for them. The NPCs’ position will remain static: even if the player waits through several of the game’s day-night cycles, the character in question will continue to wait on the arrival of the player-character, and will treat the player-character the way he/she would have if the player had chosen to drive to the mission’s starting point immediately. When the player eventually drives to that location, which triggers a brief cutscene, which would likely explicitly tell the player the nature of the mission they are about to embark on, and where they actually need to go to begin the “meat” of that mission (i.e. the car stealing, shooting, etc.) After that brief cinematic, the player would be given control of their character, instructed by some in-game text to re-enter the player-character’s car so he/she can drive to the mission proper. Given the Grand Theft Auto series’ reliance on the travel audio, there is a strong chance that another character will join the protagonist for this leg of their journey. If the car is a significant enough distance away, there is often a small exchange of dialogue between the protagonist and his new partner: a conversation that will continue as the two get into the car and drive away.

The number of this narrative modes GTA cycles through in a single mission is staggering: an extended sequence of player-action time (before the mission begins) is interrupted by a cutscene, which is followed by an episode of Potentially Offscreen Bodies, which then finally gives way to an extended bout of travel audio. Perhaps the most interesting part of this sequence is the phone call at its beginning: is it an audio-log narrative sequence, because the telephone is an acousmatic medium, or would it become travel audio if the conversation was two-way because the player-character and his phone were constantly positioned together in the center of the frame? Or perhaps both of those modes of aural-primary narrative are at work at once during
that telephone call, and the more meaningful observation would be to note how they work in tandem?

Naturally, it is the opinion of this author that acknowledging both modes is the preferred observation, as that reading of the hypothetical phone call reinforces the fluidity of the narrative genres outlined in this paper. Text-Primary, Aural-Primary, and Image-Primary are not simply all-purpose labels meant to categorize entire videogames: rather, they are also words meant to evoke certain tendencies in the transmission of narrative information in the videogame medium, which carry with them the implied effects of these tendencies’ on the player’s experience of the videogame time-space. They are terms that best categorize individual sequences within works, which gives the critic a language to speak about the content of the work with a greater level of specificity: a game is not simply “text-primary” – but a certain sequence within it might be. And it is these similarities between specific sequences that these terms highlight which allow the critic to make meaningful connections between different games without having to upend traditional videogame genre: the mode of Potentially Offscreen Bodies is vital to both a third person action game like Grand Theft Auto and a first person puzzle game like The Vanishing of Ethan Carter (The Astronauts, 2014). Giving a name to these commonalities counters the notion that the videogame medium lacks a unified language (Caldwell 42). The modes of narrative outlined in this essay establish a critical vocabulary that identifies methods of storytelling that transcend both the visual and the ergodic. These medium-specific manipulations of time and space, while greatly varied, still provide a connective tissue that solidifies the videogame as a medium with unique methods for constructing a diegetic world.

However, these methods are not meant to be prescriptive: that is, the narrative modes that are detailed in the confines of this essay do not encapsulate the full breadth of storytelling
methods one can find in the medium of the videogame. The reader may have noticed that the latter two chapters of this essay err, for the most part, on the side of the mainstream in terms of the games that have been analyzed. This was a measured choice, made to emphasize the lack of a useful critical vocabulary in videogame criticism. The rules and methods outlined in this essay can be analogized to the rules of Hollywood Continuity Editing in the film medium: that is, a set of standard practices that are synonymous with the mainstream form, which act as the foundational rules for constructing a comprehensible sense of space and time for the audience. And much in the same way as film, games that are developed outside of the mainstream system actively respond to and subvert the procedures of the blockbuster: ergo, the narrative genres outlined in this essay have been and will be subverted by multitudes of independent developers. Yet the narrative genres expounded upon in this essay remain useful. If independent/arthouse videogames can be understood as reactions to the basic operations of the blockbuster, then one must appreciate the blockbuster to be able to effectively approach the games which subvert it. For example, the aural-primary narrative is complicated by the likes of games like Deep Sea (Robin Arnott, 2011), which lack any image component, creating a diegetic space purely through sonification. Deep analysis of these subversive games is beyond the scope of this essay, but the discussion of the modern avant-garde in that concludes the first chapter demonstrates the way in which the explication of this essay’s narrative genres provides a meaningful groundwork for analyzing how art-house games reconfigure popular genres.

Such is the effect of a meaningful critical vocabulary: it provides a framework to approach a medium from that imbues the learned audience with an immediate understanding of how an object works. A critical vocabulary historicizes a medium, giving it a narrative of development and growth as innovative strategies become commonplace rules, a process which
signals the cooperative nature of a medium’s progress – how developers influence and are influenced by one another; and how critics interpret the objects and texts these developers create, granting everyone involved a deeper understanding of the nature of the medium. This essay contributes to this narrative of progress, as the modes of storytelling outlined here all identify significant, overarching tendencies within the medium as a whole, with each mode carrying with it unique implications of the player’s experience of the videogame time-space. To say that a dialogue sequence is “interactive” is to be vague and uncritical, to say that same sequence is “text-primary” is to identify the fundamental logic by which that sequence operates, and to place it within a larger tradition of videogame storytelling. For an art form so concerned with procedure, there is surprisingly little discussion of the logics by which games choose to communicate narrative information to the player. And there is ample room for further discussion: the text-primary, aural-primary, and image-primary genres of cover many of the common tendencies of videogame narrative, but they certainly do not capture the full breadth of storytelling styles offered by the videogame medium. It’ll take a great many more critics and theorists to elucidate all of the storytelling methods of the videogame. But what we’ve got here in this essay feels like an acceptable start.
Works Cited


Grimshaw, Mark. “Sound and Immersion in the First-Person Shooter.” University of Wolverhampton.


Games Cited


Crowther, William and Don Woods. *Colossal Cave Adventure*. 1976. MS-DOS.


Infocom. *Planetfall*. Infocom, 1983. MS-DOS.


Filmography