Playing with Good and Evil: Videogames and Moral Philosophy

by

Peter E. Rauch

B.A. Liberal Arts and Sciences
Florida Atlantic University, 2004

SUBMITTED TO THE PROGRAM IN COMPARATIVE MEDIA STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE IN COMPARATIVE MEDIA STUDIES
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
JUNE 2007

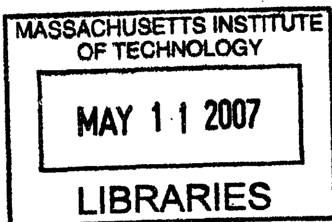
©2007 Peter E. Rauch. All rights reserved.

The author hereby grants to MIT permission to reproduce and to distribute publicly paper and electronic copies of this thesis document in whole or in part in any medium now known or hereafter created.

Signature of Author: ______________________________
Program in Comparative Media Studies
May 8, 2007

Certified by: ______________________________
Henry Jenkins
Peter de Florez Professor of Humanities
Professor of Comparative Media Studies and Literature
Co-Director, Comparative Media Studies

Accepted by: ______________________________
Henry Jenkins
Peter de Florez Professor of Humanities
Professor of Comparative Media Studies and Literature
Co-Director, Comparative Media Studies


MASSACHUSETTS INSTITUTE OF TECHNOLOGY
MAY 1 1 2007
LIBRARIES

ARCHIVES

Playing with Good and Evil: Videogames and Moral Philosophy

by

Peter E. Rauch

Submitted to the Program in Comparative Media Studies in Partial Fulfillment of the Requirements for the Degree of Master of Science in Comparative Media Studies

ABSTRACT

Despite an increasingly complex academic discourse, the videogame medium lacks an agreed-upon definition. Its relationship to previous media is somewhat unclear, and the unique attributes of the medium have not yet been fully catalogued. Drawing on theory suggesting that videogames can convey ideas, I will argue that the videogame medium is capable of modeling and critiquing elements of moral philosophy in a unique manner. To make this argument, I first address a number of questions about the proper definition of videogames, how games in general and videogames specifically convey ideas, and how games can be constructed to form arguments. Having defined my terms, I will conduct case studies on three games (*Fable, Command & Conquer: Generals*, and *The Punisher*), clarifying how the design of each could be modified to address a specific philosophical issue.

Thesis Supervisor: Henry Jenkins
Title: Professor of Comparative Media Studies

Acknowledgments

In the writing of this thesis, I am deeply indebted to several people. My three readers, Henry Jenkins, Alice Robison, and Hugo Liu, offered their considerable expertise in helping me flesh out my ideas and making sure said ideas made it onto the page in a somewhat legible form. Mark Tunick and Dan White—who constantly made me read philosophy as an undergrad even when it was clear I didn't understand it—offered valuable advice on choosing a translator for Kant. Sandra Siller helped inspire the idea of gameplay ethics during an argument about *Blood Omen: Legacy of Kain.* Finally, as always, my partner Alana offered her unwavering love and support, as did our cat, Shazzer, who always knew when I should stop writing and pay attention to her.

Table of Contents

Introduction....................7

Chapter 1: Theoretical Background and Definition of Terms....................11

Chapter 2: Kantianism and Utilitarianism in *Fable*....................32

Chapter 3: America's "War on Terror"....................63

Conclusion....................84

Works Cited....................90

Introduction

Like most in videogame studies, I've been a player longer than I've been an academic. I've been playing videogames since I was a child, and I don't remember a time when I didn't spend far more time thinking about them than actually playing them. I don't think this makes me terribly unusual in my field, nor do I think it makes the field substantially different from any other academic discipline. There's a lot of work involved in any academic field, and I don't think anyone could do it without at least some degree of personal interest in—and love for—the subject matter.

I mention this only to acknowledge that, when discussing the nature of the videogame medium, I am not a disinterested party. Outside of fan communities and academic journals, videogames are most commonly written about in terms of mass media: the economics of the industry and the effects on players. The latter is the domain of what is generally referred to as the "media effects" paradigm, a critical perspective on videogames that studies ways in which videogames might, over time, alter the behavior of players, particularly children.

I do not study media effects, and with one exception, neither do any of the theorists to whom I will refer. Nonetheless, the media effects paradigm is videogame studies' own "elephant in the room," and it would be irresponsible not to acknowledge its impact on my work. I found out about the existence of videogame studies while researching my undergraduate thesis, *Videogames as Protected Speech,* a refutation of the 2002 8th Circuit decision *IDSA v. St. Louis.* That case concerned the constitutionality of a law restricting the ability of minors to purchase M-rated videogames, a law that was passed largely in response to the claims of the media effects school. While I did not

directly address the question of whether videogames can change player behavior, I was nonetheless participating in an argument initiated by that question. My focus was on the meaning of "speech" as it had been applied to other media in previous cases, and I dealt primarily with similarities and differences between videogames and film, closing with an example of a text that clearly fit the standard for constitutional protection in a way that was consistent with precedents in film, but unique to the videogame medium in its execution.[1]

In the course of my research, a new question occurred to me. If videogames do convey ideas, they rarely convey ideas of notable substance or complexity. This is not a slight against the medium—few books or films can make the claim of communicating substantive or complex ideas, even if we eagerly protect them for that potential—but an acknowledgment that even if videogames could be constructed to make arguments, it is rarely done. Games that do make arguments exist, but they are few and far between. Of what else are videogames capable? What ideas can be articulated and defended through videogame design?

In this thesis, I have provided one answer to this question: videogames can make arguments about morality. I am not laying out a plan to improve children's behavior through videogame-based moral education, nor am I attempting to present a more pro-videogame approach to media effects. I am dealing with texts, and texts are not magical things. The legacy of the media effects paradigm is that videogames are often discussed as if they operated at an irrational, subliminal level, whereas texts of other media (such as

1 That game, *Metal Gear Solid 2,* is discussed at length in chapter 3 of "Videogames as Protected Speech." My argument will not be reiterated here.

literature) are usually treated as being rational and verbal. There are exceptions, of course—*The Communist Manifesto* and *Mein Kampf* are often described as if they have supernatural powers, overriding readers' powers of logic and free will and *forcing* them to accept the contained ideas at face value—but generally, most people seem to believe that the ideas contained in books must be voluntarily accepted through a process of rational interpretation before they can be emulated or otherwise acted upon. Videogames, when discussed on terms of their "effects," tend to be treated more like drugs. Heroin does not need to be interpreted because heroin is not a text. It goes into the user's body and does its work, whether or not the user "agrees" with the chemistry involved.

It's possible that there's some validity to this perspective, just as it's possible that there are subliminal, irrational "effects" resulting from the reading of *The Communist Manifesto,* or *Mein Kampf,* or even *The Origin of Species* or *The Bridges of Madison County*. But I am not dealing with that possibility here. Rather, I am dealing with videogames as texts that can convey ideas, in this case ideas about morality. They do so differently than other media, but I am not writing about the comparative effectiveness of media for the purposes of propaganda. It's possible that a videogame designed to argue a particular viewpoint would be quantitatively more effective, for a certain audience, than a book designed to argue the same viewpoint, but that is not what I'm writing about.

My argument is about one way in which videogames can convey ideas, and how they could do so more effectively. It is not about all videogames: in fact, my analysis applies only to single-player games with moderate linearity and a defined win condition. It is not a totalizing vision for the medium, but a new way to look at certain texts, and a new framework for articulating ideas within new texts.

In the first chapter, I will explore the theoretical background on which I base my understanding of the videogame medium, and define a number of key terms. Chief among these are "morals" and "ethics," two terms I will use throughout the thesis, delineating a structure in which videogames can make moral arguments. In the second chapter, I will examine the adventure game *Fable,* and explore how two specific moral philosophies, Kantianism and utilitarianism, could be argued in a variant of the *Fable* engine. In the third chapter, I will select two major moral issues arising from America's "War on Terror," and suggest ways to articulate moral arguments about them, drawing on a number of texts including *Command & Conquer: Generals* and *The Punisher.*

Chapter 1: Theoretical Background and Definition of Terms

In the following pages, I will use the term "videogames" repeatedly. It is common, among players and reviewers, to use "videogames" to refer only to texts played on dedicated consoles, as opposed to "computer games" played on traditional computers, but I am not making this distinction. For my current purposes, the distinction between game consoles and PCs is irrelevant. I have decided on the term "videogames," as opposed to "video games" or "computer games," because "videogames" suggests a category unto itself, rather than a type of game that merely *happens* to involve the use of a video screen or personal computer. In addition, "videogame theory" is less likely to be confused with "game theory."

Videogames, as a term, presents two etymological problems. First, it provides no easy distinction between the medium and individual texts, as can be inferred from terms like "film" and "films." It is not always aesthetically feasible to use the phrase "the videogame medium" whenever I refer to the medium as a whole. Second, while technically only one word, "videogames" remains a composite. All commonly used terms for the medium are composites that add a media-indicative adjective such as "video" or "computer" to the pre-existing noun "games." This relegates the medium to a subcategory of the existing cultural understanding of games. While games and videogames have much in common, the connection is reinforced so strongly in language that perspectives on videogames' aesthetics and capabilities are somewhat limited, as has been demonstrated in both academic and popular discourse. This latter problem confuses not only issues of vocabulary but the important distinctions between games in general and videogames. To explore these distinctions, I will examine existing theory concerning

both, and delineate my own provisional definition for the medium. Nonetheless, when I refer to "games," I am generally speaking of videogames, unless context indicates otherwise.

Johan Huizinga, one of the earliest theorists to systematically study play, argues that games are a fundamental part of human nature, and that the urge to play underlies everything from the simplest childhood game to the most reverent religious ritual. Play is so fundamental, he argues, that it *precedes* humanity:

> Play is older than culture, for culture, however inadequately defined, always presupposes human society, and animals have not waited for man to teach them their playing. We can safely assert, even, that human civilization has added no essential feature to the general idea of play. Animals play just like men." (1)

This is a crucial idea, and the starting point for thinking about games, ethics and morality: play, far from trivial, is a serious matter. Huizinga claims that, in games, the rules are absolute and immutable. "The rules of a game are absolutely binding and allow no doubt. [...] Indeed, as soon as the rules are transgressed the whole play-world collapses. The game is over" (Huizinga 11). To obey the rules is to play the game, and to ignore the rules is to break the "magic circle" that demarcates play space from reality (Huizinga 11). While the rules may not be moral—they are, after all, removed from standard reality—they are absolutely imperative within the magic circle, because the circle cannot be sustained without them.

Expanding on Huizinga's work, Roger Caillois makes a distinction between two types of play, *paidia* and *ludus*. He defines *paidia* as the "basic freedom [...] central to play in order to stimulate distraction and fantasy," and *ludus* as "the taste for gratuitous difficulty" that manifests in the forms of specific rule systems (141). *Paidia* represents

unstructured play, as manifested in "a cat entangled in a ball of wool, a dog sniffing, [or] an infant laughing at his rattle" (141). *Ludus,* on the other hand, represents more structured play, the activities commonly referred to in English as "games." *Paidia* and *ludus* might be better viewed as extrema on a continuum rather than two discrete categories. Videogames like *Second Life,* in which *rules* are present in the sense that the game is built on specific code and requires a specific user interface, but there exists no clearly defined *goal,* seem to fit neatly into neither category. However, the texts I'll be discussing heavily favor *ludus* over *paidia,* encouraging (perhaps forcing) players to play *in a certain way.*

The way in which the game is to be played is "taught" to the player through positive and negative feedback: players experiment with an environment that rewards correct actions and punishes incorrect ones. Raph Koster takes a cognitive approach to games, and suggests that games are learning machines, systems of puzzles designed to provoke certain reactions and impart certain tactical knowledge (36). He claims that, so far, most videogames have been teaching simplistic ideas about power and dominance: "Consider the games that get all the attention lately: shooters, fighting games, and war games. They are not subtle about their love of power. [...] They are all about reaction times, tactical awareness, assessing the weaknesses of an opponent, and judging when to strike" (58). While existing games seem to teach primarily about acquiring and exercising power, Koster suggests that greater creativity on the part of the designers could lead to more complex, socially relevant ideas in the future (68). He is somewhat conflicted about the role of narrative in games, however, and does not seem to grasp that ideas like power and dominance cannot be conveyed to players through completely

abstract symbols; his treatment of "dressing," the elements of game design not related to rule systems, is deeply ambivalent about the importance of aesthetics or narrative (168-170).

Ironically, Koster shares many ideas with vocal videogame critic Lt. Col. David Grossman, a retired military psychologist who claims that many currently popular videogame genres are derived from military exercises designed to break down the natural inhibition to kill. In the military, he argues, these exercises train soldiers to be brutal killers on the battlefield, able to kill even when their "forebrain" has shut down and the resistance to killing should be strongest (Grossman, Trained to Kill). He goes on to suggest that this process led to both the high fire rates of soldiers in Vietnam, as well as the high rate of post-traumatic stress disorder (PTSD) in returning veterans. Yet, even veterans suffering from PTSD are less likely to commit crimes than non-veterans. "The key safeguard in this process appears to be the deeply ingrained discipline which the soldier internalizes with his military training" (Grossman, On Killing II). This discipline, however, is not provided by videogames, which teach only raw skills. Grossman claims that military discipline offers a moral context that clarifies "good" and "bad" violence, but videogames cannot generate such context (because they are entertainment), and teach that all violence is fun and inconsequential. While Koster seems to believe that the skills taught in the modern high-profile games he describes above are more or less unrelated to the game's "dressing," Grossman suggests that the dressing is crucial, and the nature of entertainment causes unconscious mimicry by players. Koster says that the moral value of shooting a gun in a videogame is neutral, while Grossman says it is negative; they agree, however, that it cannot be positive, and that moral context is not a province of

game design. Grossman's thoughts on how moral context is created are at times contradictory, but the idea that moral context can only come from outside the game is a popular one even among gamers, and one I intend to challenge.

Espen Aarseth helped define the field of ludology, a branch of videogame theory that downplays the importance of narrative in favor of player-directed action, i.e. game mechanics. He also first applied to videogames the term "ergodic," a term from physics meaning "path-dependent." "Ergodic" represents a higher level of reader/play involvement than "interactive," since the former involves active, exclusionary choices and the latter involves many activities, some as subtle as mere interpretation. "In ergodic literature," e.g. multi-user dungeons (MUDs), "nontrivial effort is required to allow the reader to traverse the text" (Aarseth 1). While effort is required to *interpret* any literary text, *traversing* a literary text requires only a movement of eyes across a page, which must periodically be turned. Aarseth elaborates:

> A reader, however strongly engaged in the unfolding of a narrative, is powerless. [...] He cannot have the player's pleasure of influence: "Let's see what happens when I do *this.*" The reader's pleasure is the pleasure of the voyeur. Safe, but impotent. [...] The cybertext reader, on the other hand, is not safe, and therefore, it can be argued, she is not a reader. The cybertext puts its would-be reader at risk: the risk of rejection. The effort and energy demanded by the cybertext of its reader raise the stakes of interpretation to those of intervention. (4)

Ludologists, like Aarseth, reject attempts to interpret videogames using tools derived from narratology, preferring to conceive of videogames as cybernetic systems or simulations. In Simulation *versus Narrative: Introduction to Ludology,* Gonzalo Frasca designs a hypothetical game called *Strikeman.* Drawing inspiration from Emile Zola's *Germinal* and Ken Loach's *Bread and Roses, Strikeman* is videogame that simulates a

worker's strike. Noting the predominantly negative ending in Zola's novel and the predominantly positive ending in Loach's film, Frasca argues that the authors of each are limited by the representational nature of their chosen media. "Both storytellers are arguing that change is possible. However, neither of them is telling us to what degree that change is possible" ("Simultation" 226). He continues:

> Narrative authors or "narrauthors" only have one shot in their gun—a fixed sequence of events. At most, they could write five or six different stories describing strikes, so the reader could make an average and decide the probabilities that workers have to succeed. [...] In a game, going through several sessions is not only a possibility but a requirement of the medium. Games are not isolated experiences: we recognize them as games because we know we can always start over. ("Simulation" 227)

Because videogames are inherently iterative, they can produce infinite variations on a single event, and this gives them the ability to communicate about these events in a unique manner. In *Strikeman,* players take on the role of a labor organizer attempting to organize and implement a strike. Every play session would necessarily be slightly different: "[i]n one session, the boss could call the police and repress your workers. In another game, you may have to deal with spies infiltrated into your organization or another worker may contest your leadership and try to sabotage your actions" ("Simulation" 227). In any game, the strike's success would depend both on the player's skill and the interaction of random elements. Much of any videogame is completely unrelated to the player's actions, and even in the most non-linear game, the designer maintains authorial control by deciding "the frequency and degree of events that are beyond the player's control" ("Simulation" 228). Through careful, deliberate rule design, *Strikeman* could make an argument about worker's strikes that is qualitatively different than the kind made possible by a representational (as opposed to simulational) medium,

incorporating elements of probability, chance, and inevitability. In addition, all of this information will be implicit in the game's rule system, to be discovered by the player through experience, never without the proper context. "Narrative may excel at taking snapshots at particular events," writes Frasca, "but simulation provides us with a rhetorical tool for understanding the big picture" ("Simulation" 228).

Henry Jenkins has made some effort to reconcile what he perceives as a false and limiting binary between narrative and simulation by suggesting a view of games as narrative architecture, telling stories and conveying ideas through the traversal of virtual spaces. "Game designers don't simply tell stories; they design worlds and sculpt spaces," he writes, placing the videogame medium within a spatial storytelling tradition that includes both paper mazes and Japanese scroll paintings ("Narrative Architecture," 674-675). Jenkins proposes several ways in which stories are enacted in game spaces, including embedded narratives and emergent narratives. Embedded narrative consists of a story that is not explicitly told to the player, but rather embedded in the game's architecture in a way that can be inferred by an attentive player; Jenkins cites *Myst* and *Half-Life* as prime examples ("Narrative Architecture," 682). Emergent narrative, on the other hand, consists of narratives that are not explicitly scripted by designers, but are also "not as unstructured, chaotic, and frustrating as life itself" ("Narrative Architecture," 684). Both embedded and emergent narratives are consistent with Frasca's vision of *Strikeman,* but Frasca does not acknowledge the extent to which some level of narrative is a necessary precondition for a simulation, and not merely an "option" for ambitious or pretentious game designers. A simulation must, by definition, simulate a thing, and players' associations with the simulated are inevitably carried over into the simulation,

creating an "evocative space" ("Narrative Architecture," 677).

Stephen Poole seizes upon the relationship between simulation and simulated, describing videogames as "semiotic engines," machines that generate play experiences through the interplay of symbols (214). Discussing the role of narrative in videogames, Poole makes a crucial distinction between "synchronic" and "diachronic" narrative: "For the purposes of talking about videogames, the 'backstory' is the diachronic story, and the story that happens in the fictional present is the synchronic story—an ongoing narrative constituted by the player's actions and decisions in real-time" (106). Nearly all stories, in all media, involve both elements: the synchronic narrative being what occurs in the "present," and the diachronic narrative being the events that precede that artificial present. Since any ergodic elements must, by their very nature, take place in the present, only the synchronic can be represented through gameplay, while the diachronic is explored almost entirely through non-ergodic means, such as cut-scenes, dialogue windows or prose exposition. Poole notes that most narrative complexity, in the form of dialogue, character development and major plot progression, occurs in the diachronic, while the synchronic tends to deal almost exclusively with actions such as travel or combat (107-108). While the two are not so clearly defined in every game, the model is quite useful, and the synchronic/diachronic binary is highly reminiscent of the simulation/narrative binary. Poole suggests that interactive narrative in videogames, the holy grail of interactive storytelling, is not yet possible, arguing that "[n]ot only has no convincing example of this new creature called 'interactive storytelling' yet been spotted in the wild, no one is even sure what it might look like" (106). Poole's argument against the possibility of interactive storytelling certainly seems convincing, but only because it

is predicated on an unnecessarily restrictive view of the medium. Every videogame contains portions in which the player cannot affect the game world through the control apparatus, but these non-ergodic moments are still part of the play experience, providing context and meaning for the gameplay.[2]

Context and meaning are crucial to James Paul Gee's understanding of the videogame medium. Gee argues that videogames' ability to model worldviews, or "cultural models," allows players to articulate and challenge their own unexamined assumptions about the world (143). Gee examines a variety of war-themed games, from the superheroic *Return to Castle Wolfenstein* to the darkly realistic *Operation Flashpoint* to the explicitly political *Under Ash*. *Under Ash* (an action game in which the player takes on the role of a Palestinian fighting against Israeli soldiers and settlers), and hints at an unrealized potential of the videogame medium:

> One of the things that makes video games so powerful is their ability to create whole worlds and invite players to take on various identities within them. When players do this, two things can happen: On one hand, their presupposed perspectives on the world might be reinforced. [...] On the other hand, through their creation of new and different worlds and characters, video games can challenge players' taken-for-granted views about the world. (139-140)

All of these theorists have contributed to my understanding of the videogame medium, in terms of what it is, what it can do, and what it can help to inspire players to do. Huizinga, writing before the existence of videogames, suggests that games are systems of rules that, while not apparently important to normal life, are absolutely crucial and non-negotiable in the context of the game itself. Caillois expands on this, suggesting

2 In "Hands-On Horror," Tanya Krzywinska argues that certain genres (such as survival horror) actually rely on the tension between active and passive gameplay, and couldn't function without it. I explored these ideas in relation to representations of moral ambiguity in "Free Will in a Predestined World: *The Suffering* as Interpretive Space."

a more complex typology of play and games that varies on the balance between strict rules and improvisation. Koster suggests that games have a practical effect, teaching players raw skills that can be applied to other areas of life, although he notes that certain skills are more applicable than others to a given historical moment. Grossman, whose theory I eschew on its own merits, serves to highlight what I consider to be a blind spot in Koster's work, the idea that games deal only with skills and that the context for applying those skills can only come from outside the game. Aarseth places videogames in the context of previous media (specifically, literature), and clarifies that videogames present an experience that is qualitatively different from previous media, no matter how similar they may superficially appear to be. Frasca classifies videogames as simulations, in which designers encode a potentiality of narratives to be enacted by players. Jenkins articulates the importance of spatiality in videogame narratives, both as a means to suggest implicit narratives or to draw extra-textual narratives into the play experience. Poole emphasizes the importance of temporality in videogame narratives, demonstrating that active and passive gameplay lend themselves to different kinds of narrative. Finally, Gee suggests that videogames can force players to assume perspectives other than their own, and challenge their internal assumptions about the world.

Drawing on these theorists, my provisional definition of "videogame" is thus: a game—that is, an abstract rule system—realized and enforced primarily through code interpreted by a computer, that is semiotically mapped to a narrative and represented visually or audiovisually, and designed to create a marginally variable experience for the player. Videogames can tell stories, but more fundamentally they can convey ideas.

Many theorists and designers have already begun to explore this communicative

potential of videogames. In *Videogames of the Oppressed,* Gonzalo Frasca suggests integrating ludic principles with Augusto Boal's participative drama techniques in order to foster critical thinking skills in players and player communities. Frasca also designs hypothetical games, as I will do in later chapters: one derived from *The Sims,* and one derived from several "videogame primitives" such as *Pac-Man, Tetris,* and *Street Fighter* (*Videogames* 95). In addition, the "serious games" movement seeks to harness the power of videogames for education, training, and various public policy issues.[3]

What I am suggesting here is slightly outside the area marked out by Frasca and the serious games movement. My primary goal in this piece is to establish a set of fundamental principles for dealing with moral arguments in games, "serious" or otherwise. The case studies that follow are a means to test and refine those principles, and I have selected games with strong *ludus* rules from popular commercial genres. While Frasca chose his list of "primitives" because these early games "were so technologically constrained that they had to focus on the essence of the action" (*Videogames* 95), I have built my case studies on more recent games with the technological freedom to explore more detailed narratives. While moral philosophy is certainly a subject associated with education, it is not one with which educational games have traditionally been associated, being a subject that is traditionally studied in college curricula or not at all. While I focus on the ability of games to make arguments, I am not so concerned with their ability to convince players as the usefulness of having videogame texts with a clear point of view that is articulated by means of the medium's unique

3 To this end, Frasca designed *September 12th*, a game about the "War on Terror" that will be discussed in chapter 3.

properties; I am not a game designer, and I am here concerning myself with the "how," not the "what." To delineate these fundamental principles, a vital distinction must be made between morals and ethics.

Many dictionaries consider morals and ethics to be synonymous, but in common usage, at least in American English, the two words can have a variety of subtly different meanings. My definitions are provisional, and while they can be used in general discussion, they are specifically tailored to be applied to the interpretation of videogames.

I define ethics as a discourse concerning what is correct and what is incorrect. What is ethical is dependent on a specific activity, determined entirely by an explicit, constructed system of rules, and cannot be questioned by the participants. I define morals as a discourse concerning what is right and what is wrong. Morality, unlike ethics, is not tied to a specific activity, but can be applied over multiple activities, and possibly all experience. Moral rules enjoy considerably more variance than ethical rules: because they are wider in scope, they are more nuanced, and subject to interpretation.

Ethical frameworks, while they might attempt to model moral behavior—as in the examples of ethical codes for doctors or lawyers—need not have any connection to morality at all. In chess, that players should try to capture their opponents' pieces is an ethical rule, not a moral one. It has no relevance to the world outside chess. This rule is also not subject to interpretation or argument. It is simply, factually, true. A player that makes no effort to capture the opponent's pieces is not playing chess. The same cannot be said of moral rules like "love your neighbor as yourself," Jesus' formulation of the "golden rule," nor can it be said of "act only in accordance with that maxim through which you can at the same time will that it become a universal law," Kant's categorical

imperative. These rules concern the very act of being human, but one does not cease to be human if he or she rejects or violates them. They are much less specific than the rule concerning the capturing of pieces in chess, and open to many more interpretations.

Morals and ethics exist independently of each other, and while they must each be internally consistent, it is possible for the two to explicitly contradict one another. Law is an ethical system that is constantly revised to prevent such conflicts. Torture, for example, is illegal under international law. Assuming one accepts the existence of international law, the legality of torture is not open to debate. The morality of torture, however, is fundamentally unconnected to its legality. Torture is not less moral now than it was before the Geneva Convention. Conversely, it would not become more moral if the U.N. were to repudiate the Geneva Convention tomorrow.

Any game that has a "win condition" has an ethical framework. This applies to all games, not just videogames. First and foremost, these games are possessed of an overriding ethical imperative: win. If the game has a win condition, a player who does not try to win is not playing the game. As Johann Huizinga notes in *Homo Ludens,* a player who does not try to win faces greater censure from society than a player who cheats in order to win (11). One interpretation of Huizinga's claim is that a player who cheats breaks only those rules concerning the *means* of play, whereas the player who throws the game violates the *goals* of play. The goal constitutes what players must do, while the rules offer only clarification on how the goal is to be accomplished—what actions are correct, and what actions are incorrect. A strategy or technique that helps a player win, while not explicitly violating any of the rules, is always ethical, in terms of the game in question. The ethical framework comprises both goal and means, and

although the former is more fundamental to the game than the latter, they are both necessary for a game to function. The game's ethics, which determine how it *should* be played, are inextricably bound in the game's rules, which determine how it *can* be played.

I use the term "ethical" to denote imperatives that are dependent on the accepting of a role, as in the specific ethics of a given profession, and also in terms of play in general—playing a videogame ethically could be seen as the player's agreement to play the role allotted to her by the designers. Some degree of freedom is present, of course; were such freedom absent, it would not be play. However, just as an actor may be allowed to improvise, but must ultimately play his role to the author's conclusion, the player must play "in character" to play the game. If the player does not accept this role, she is not playing the game, but rather playing a game *with* a game. This activity of "metaplay,"[4] in which the player designates goals unrelated or contrary to the game's internal ethics, has a wide variety of forms, some showing clear principles of *ludus* (adding or removing barriers to make the game easier or more difficult), others showing behaviors associated with *paidia* (exploration for its own sake), and still others being more difficult to determine (making machinima). Metaplay, at least in single-player games (where there are no social expectations of ethical play), is not "cheating" in the sense that the word is used in everyday speech. It simply means that the player in question is not, strictly speaking, playing the game.

In addition to the ethical frameworks inherent in any games, videogames can potentially add an unprecedented level of narrativity. This narrativity is achieved by

4 This term is not synonymous with "metagame," the term Henry Jenkins uses to describe the social context in which games are played ("Effects and Meanings" 214)

mapping recognizable symbols onto the rule system. This mapping process allows for the suspension of disbelief necessary to involve the player emotionally in the gameworld.

The interaction of these symbols gives videogames the potential for rich narratives. However, if the narrative is not sufficiently integrated with the rule system, it will appear arbitrary, and fundamentally disconnected from the experience of play. This disconnect between narrative and rule systems is one of the central problems for the potential of videogames as a communicative medium, forcing a distinction between authorial narrative (the story written by the designers) and emergent narrative (the story enacted by the players). Even in the most non-linear games with the greatest potential for emergent narrative, the rule system and choice of symbols are selected by the designers, and as such the players' freedom of interpretation is inherently limited. In videogames, the author might be dead, as was famously suggested by Roland Barthes, but she is still the author, and she must not be confused with the reader. To make the transition from ethical imperatives to moral argument, the designers must fully embrace authorial status.

Narrative alone is not sufficient for morality, of course, since without a connection to the ethics, the gameplay and the narrative will operate independently of one another, as is often the case in games that rely extensively on cut-scenes. Moral imperatives can exist in a game only when the ethics can be interpreted and applied to the world in which the game is played, and this can only be achieved by connecting internal ethics to the external world through narrative. Most, if not all, of the game rules must be connected to recognizable symbols, and those symbols must have referents in reality.

Rules and a win condition are all that is necessary for an ethical framework, because ethics point inward to a specific activity. Conversely, because morality must

gesture outward to the world at large, it cannot consist only of abstract symbols. For a game to have a moral framework, it must have an ethical framework, a narrative that can be connected in some way to what we speciously refer to as "real life," and a careful integration of the two. Specifically, the moral argument of the narrative must be connected to the win condition. It might be necessary, in making distinctions between what is right and what is expedient, to develop some new ideas as to what constitutes "winning." This will require a somewhat nuanced perspective on the avatar.

The avatar, in most games, is more than an extension of the player into the gameworld. Rather, the avatar is simultaneously an extension of the player and a different character that is *not* the player. I refer to this different character as the protagonist. Since the protagonist has only diegetic information, his motivation for interaction in the world must be entirely diegetic. The player, who has access to the game's non-diegetic information, will have additional goals, often involving tasks with no narrative meaning such as scoring points or unlocking content. Narratives, even videogame narratives, have a logic of their own, and even when the narrative fails to emotionally invest the player in the story, it can usually be assumed that the protagonist is quite involved. The narrative, even when viewed by players as epiphenomenal, is the entirety of the protagonist's reality.

These constructed realities often have moral rules of their own, and videogames allow players to inhabit characters of any conceivable moral persuasion. In most games, the protagonist or protagonists are ostensibly "good," for reasons that are too numerous and complex to be sufficiently explained here. When players are offered the chance to inhabit a character who is ostensibly "evil," a complicated semiotic exercise occurs. Gee

describes this phenomenon in reference to Shadow, a villainous protagonist of *Sonic Adventure 2 Battle*:

> If you want to play *Sonic Adventure 2 Battle* from Shadow's perspective you must act, think, and value (while playing) from [...] a perspective that makes Shadow "good" or "the hero." After all, you are fighting numerous battles as Shadow and feel delight when winning them and dismay when losing them. It would be absolutely pointless to play as Shadow but purposely lose battles because you disapprove of his value system. If you played that way, Shadow would die quickly in the first episode and you'd never see anything else in the Shadow part of the game. (142)

Sonic Adventure 2 Battle is one of many games that allow players to choose from multiple moral paths within a single world. Were the player to actually exist in that world, it would almost certainly be in his or her best interest to "root" for Sonic, the "good guy," over Shadow. These distinctions are generally painted in broad strokes, in which both good and evil are portrayed in cartoonish extremes. Other games allow players to inhabit protagonists who are morally undefined, and can choose either good or evil. Sometimes this choice is made through very specific events, usually at the end of the game (as in *Streets of Rage* and *Blood Omen: Legacy of Kain*). More often, the games make use of a moral model to decide the protagonist's morality by the sum of many different actions.

Moral models are common enough in videogames, but have generally been fairly superficial. The most common method is a simple axis in which "good" actions push the player's moral "rating" in one direction, and "evil" actions push it in another. It is a simple model, to be sure, but nonetheless it presents problems. In some games, the actions available to the player are so obviously moral or immoral that the game can make no legitimate argument about morality. This is not specifically a problem, since it is

unlikely the games in question were actually *intended* for such a purpose. A more substantial problem exists if the difference between good and evil is largely superficial. One particularly egregious example exists in *Star Wars: Dark Forces II: Jedi Knight.*

In *Jedi Knight,* the protagonist is Kyle Katarn, a mercenary who trains as a Jedi and learns the ways of the force. As players progress, they gain access to certain "force powers," some morally neutral, some associated with the "light side," and some associated with the "dark side."[5] The largest influence on the protagonist's morality is the choice of force powers. How a given power is determined to be "light" or "dark" seems rather arbitrary, since every force power has the ultimate effect of making the protagonist a more effective killer. However, a larger problem exists with the mundane application of morality. Before force powers are an issue, Katarn's morality is affected only by his actions during the course of gameplay. The only evil action available to the player is the killing of unarmed non-player characters (NPCs). The only good action, it would appear, consists of *not* killing unarmed NPCs. Killing these NPCs usually provides no benefit whatsoever to the player, so essentially a "good" protagonist is one who doesn't kill defenseless people purely for the hell of it. In terms of modeling morality, I feel this to be setting the bar a little low.

Other games use slightly more complex moral models. *Neverwinter Nights,* drawing on the *Advanced Dungeons & Dragons* ruleset, gives players nine moral identities, derived from two axes: good/evil, and lawful/chaotic. This allows for slightly more complex character development, but ultimately just invites even more questions as

5 The term "light side" appears in none of the *Star Wars* films, and it bothers me intensely, but it seems to be the only way to mark this binary. But I digress.

to what constitutes good, and what constitutes law.[6] Another example of a more ambitious moral model is *Black & White.*

The creation of longtime strategy game designer Peter Molyneux, *Black & White* casts the player as a deity in a polytheistic universe in which gods compete with one another for power and influence. Players are, of course, allowed to pursue good and evil at their extremes, but both extreme moral paths present different challenges. Benevolent gods who help worshippers with their day-to-day survival can expect thriving populations, but also less self-sufficient worshippers. Inspiring worship with violence and terror will deplete the pool of available worshippers quite quickly, especially when coupled with a laissez-faire approach to human survival. In *Black & White,* a god's power is finite and tied to the number of worshippers and the intensity of their worship, and evil—usually the "easy" path in games that provide players the choice—carries complications of its own. For good or for ill, any god's actions inspire less awe the more often they are repeated, leading to a law of diminishing returns on miraculous behavior in general. Combined with the personalities of the worshippers themselves, the player must perform a delicate balancing act:

> [A]s the game goes on, [the worshippers'] needs expand, forcing the player to micromanage his or her everyday wants ever more attentively. According to Richard Evans, the game's artificial intelligence programmer —who based much of it on concepts he learned while he was a philosophy student at Cambridge—this was entirely intended by Peter Molyneux [...] "Peter wanted the villagers to be increasingly reliant [...] so that the more you helped them, the less self-sufficient they are, so that you are drawn into a spiral of dependency. He was trying to make a point about human

6 Games drawn from pen-and-paper role-playing games tend to reflect their heritage through a greater emphasis on character statistics, and often representations of morality follow suit. The 1986 text-based computer game *Alter Ego* allows players to experience a virtual human lifespan, and arguably contains its own moral engine that does not easily fit existing categories. An emulation of *Alter Ego* can now be found on the web.

nature." (Au)

Despite the binary implied by its title, and the general "helping good, hurting bad" rule that seems to govern the game's "rating" of the avatar's morality (represented both numerically and graphically), *Black & White* seems designed to draw the player into a more nuanced morality. It is possible to make a coherent philosophical argument that the "evil" path is more moral than the "good," or that a synthesis of the two is ideal. Hobbes and Machiavelli would certainly find more to admire in a player who dwelled in the middle than one who rendered herself powerless and unable to provide for her dependent worshippers through excessive mercy. What *Black & White* adds to the moral model is a philosophical space in which nuances can be brought into focus and debated. The philosophy in question can be difficult to apply, since the morality of a god, even a finite god, is difficult to compare to the morality of a human. The closest real-world analogue to the player's activity in *Black & White* would be (rather monarchic) statecraft, not individual behavior. This might be why Molyneux's next game, *Fable* attempted to fashion a similarly complex moral model at the individual human level. *Fable* is the subject of the next chapter.

The relations of the rules to the narrative, the player to the protagonist, and the ethics to the morals are key to videogames' potential to model and critique elements of moral philosophy. In the following chapter, I will explore hypothetical games that argue the validity of the Kantian and utilitarian moral perspectives. These two perspectives represent only a small portion of an enormous field, of course, ranging from the deontological paradigm (deriving morality from the observance of rules or laws), to the consequentialist paradigm (deriving morality from the consequences of actions), to the

aretaic paradigm (finding morality in a holistic balance of virtues). Kant's moral philosophy is deontological, whereas utilitarianism is consequentialist. I have chosen to focus on Kantian and utilitarian philosophies for several reasons. First, they directly oppose each other in many ways that could be easily represented in existing videogame genres. Commercial videogames tend to hew toward an intuitive morality that adheres to no particular school, but could be seen as a synthesis of several mutually exclusive moral philosophies. The value of designing a game around an existing philosophical perspective is greatest when that perspective is placed in conflict with other moral perspectives. Games based on Kant and utilitarianism will demonstrate these points of conflict in many of the same places. Comparing the two alternate models of *Fable* makes the structural differences between the two more apparent. In addition, the texts I'll be working with are chronologically fairly close together: Immanuel Kant published *Groundwork of the Metaphysics of Morals* in 1785, while Jeremy Bentham published *Introduction to Principles of Morals and Legislation* in 1789, with John Stuart Mill expanding on Bentham's work in *Utilitarianism* in 1863. By placing these nearly contemporary, but directly opposed, philosophies next to each other in hypothetical videogame form, I will explore how the proposed morality influences even the most fundamental rules of gameplay. I will examine Kantian and utilitarian adaptations of *Fable* in turn, and conclude by comparing and contrasting the two models.

Chapter 2: Kantianism and Utilitarianism in *Fable*

"For Every Choice, A Consequence," reads the ad copy for *Fable.* It is a bold claim, from a design perspective, hinting at the holy grail of interactive storytelling so derided by Poole. A third-person adventure game, *Fable* occupies a well-worn genre. Its claim to originality developed from its treatment of morality. Like *Black & White, Fable* received a great deal of press during its long development time, and also like *Black & White,* it was perceived by many players that the creators' ambitious promises were not realized in the final game design. In practice, *Fable* is an unremarkable adventure game, and while many actions do have moral consequences, these consequences are predominantly superficial. However, a great potential for moral argument remains inherent in the design.

Fable is a world to be explored, with an ethical framework to be discovered through play. The narrative involves human-like characters that can lie, coerce and kill each other. Some of the actions of these characters are categorized as "good" or "evil" according to the beliefs of the designers, and the player is relatively unrestricted in choosing to perform them. Finally, the player's actions are acknowledged by the rest of the world, however imperfectly, in the sense that NPCs respond to the player's past actions, as well as the avatar's appearance. *Fable* possesses all the raw materials to create a convincing, semi-realistic world that is intentionally biased toward a specific worldview —to argue the validity of a moral philosophy. That this possibility was not realized, or not sufficiently realized, or not meaningfully realized, does not alter the game's potential. As such, *Fable* seems an ideal place to start when conceptualizing games that make

meaningful arguments about morality.[7]

To this end, after exploring *Fable* in some detail, I will investigate two strictly opposed moral philosophies, Kantianism and utilitarianism, and suggest ways in which *Fable*'s design could be altered to argue these viewpoints. I will then examine specific points on which the two philosophies are opposed, such as honesty and justice. Finally, I will suggest ways in which, by opposing the imperatives of gameplay and narrative, satirical arguments could be made.

The player begins *Fable* as a (male) child in a small, fantasy-medieval village in the land of Albion. Childhood functions, rather appropriately, as a tutorial, introducing the player to most of the basic play mechanics, as well as the game's moral engine and social system On the day in which the game begins, it is the protagonist's sister's birthday, and he needs money to buy her a gift. His father, eager to cultivate noble habits in the boy, offers the protagonist a coin for every good deed he does. The player is then presented with several conflicts demanding his or her intervention: each allows the player to make right or wrong choices, and the player is explicitly told the morality of his or her choices by a change in the protagonist's "alignment." The player can engage in these conflicts in any order; I have numbered them here only for convenience.

In the first conflict, a little girl tells the protagonist that her teddy bear has been taken. Elsewhere in town, the protagonist finds a little boy being threatened by a bully. The little boy is in possession of the teddy bear in question. The bully, the player learns through dialogue, is the little girl's older brother, and wants the teddy bear so he can

7 My examination of *Fable* is based on *Fable: The Lost Chapters* for Windows. To my knowledge, the elements I discuss are common to all iterations of *Fable.*

destroy it. (How the little boy came to be in possession of the teddy bear in the first place, in such a way that its owner was unaware, is never fully explained.) The bully offers to pay the protagonist one coin to get the teddy bear from the little boy. Here, the player has two initial options: he or she can beat up the bully, or pummel his victim. If the player chooses the former, the bully will begin whining with the first blow, and eventually run away. In this case, the little boy thanks the protagonist and gives him the teddy bear, which can then be returned to the little girl. Both assaulting the bully and returning the bear to its owner are considered "good," and have a positive effect on the avatar's alignment. If the player chooses to assault the little boy instead, the boy will complain about this injustice and give the player the teddy bear in an attempt to stop the violence directed at him. At this point, the player faces another choice: to give the teddy bear to the bully, receiving a coin as reward, or take the teddy bear to the little girl, performing a good deed for which the protagonist's father with also pay him one coin. Attacking the boy is a "bad" action, as is giving the teddy bear to the bully—each gives the player two "evil" alignment points. Returning the bear to its owner is a "good" action, worth two "good" alignment points. Consequently, a player who hits the little boy and then returns the teddy bear to its owner will end up with the good and bad actions cancelling each other out, numerically, although the player can get an additional two "good" points by attacking the bully after the fact. No matter which course of actions the player chooses, the protagonist will end up with one coin.

In the second conflict, a woman complains of her philandering husband, and asks the protagonist to find out where he is and what he's doing. Sure enough, the player finds the man engaged in an amorous embrace with another woman—upon discovery, he offers

the protagonist a coin to keep quiet. (The game warns the player that rumors travel fast in the village, and people will know he took the bribe.) If the player takes the bribe, he or she receives two "evil" points and gets the coin, although he or she can balance those points out by breaking his promise to the adulterous husband and telling his wife the truth. From a monetary perspective, this is the ideal solution, since the player gains two coins, one from the husband and one from the protagonist's father for doing a good deed.

In the third conflict, a merchant asks the protagonist to watch his barrels while he runs an errand in town. Some local boys urge the protagonist to break them and see what's inside. Honoring the merchant's wishes gets the player two "good" alignment points and a coin from the protagonist's father, while smashing all the barrels earns the player two "evil" alignment points and a coin from inside one of the barrels. Curiously, if the player can break all the barrels and get back to where the protagonist was supposed to be standing guard before the merchant returns, the merchant will thank him for watching the merchandise, and the player will receive two "good" alignment points and a coin from the protagonist's father, despite having broken his promise. Again, the "neutral" path, i.e. performing both good and evil deeds with no apparent logic connecting them, presents the fastest way to earn money, buy a gift for the protagonist's sister, and advance in the game.

At first glance, it would seem that these examples do not lend themselves to moral subtlety. Even in "real life," morality is taught to children first in broad strokes, and the morality of many fantasy worlds is similarly rendered in black-and-white. It makes perfect sense, from a design perspective, to deal with morality on a very simple level in the tutorial and flesh it out as the game continues. However, the *Fable* tutorial fails to

accomplish even this, because of a poorly thought-out reward system that severely limits players' choice of action and defines morality in terms of discrete actions, regardless of motive or intent. In the conflict involving the teddy bear, no non-violent options exist: the player cannot attempt to reason with the bully or threaten him verbally. While it can be argued that some conflicts can only be solved through the judicious application of violence—*Fable* is an adventure game, after all, and much of the game is spent killing—few would argue that this is *necessarily* the case for conflicts involving children, and that beating up the bully is the *best* moral option available to the player. In addition, the "evil" alignment points given to the player for hitting the little boy can be cancelled out by attacking the bully, even though there is no logical reason to do so. Therefore, in *Fable,* random, illogical violence for the sake of violence is perceived as morally superior to violence as a means to an immoral end. Similarly, it could be argued that taking the adulterous husband's bribe and then telling his wife anyway is, morally, the *worst* option, since it could be interpreted to represent an amoral pursuit of profit. Finally, that the player can break the merchant's barrels without him realizing it, and be rewarded for it, simply makes very little sense.

The problems presented here are twofold. First, it seems that *Fable*'s designers put very little effort into deciding *why* given actions are right or wrong. Actions are decided to be moral or immoral, but few clear principles seem to have been defined to guide these decisions, and those that do are not consistently applied. Second, the game as it currently exists can respond to play actions, but not player intent. The importance of intent in morality is hotly debated of course, and intent *is* coded into *Fable* by the designers. However, the player has no role in deciding this intent. Like the protagonists

of many adventure games, *Fable*'s hero is presented as a *tabula rasa,* and the player never hears him speak (dialogue choices are generally presented as a simple "yes" or "no"). A character who cannot speak cannot easily articulate his intent, but this intent does nonetheless exist at a narrative level. If the player's intent for a given action differs from that which the designers scripted, the result can be quite jarring.

Some of my own experiences can attest to this. Once, early in the game, I was called upon to to attack an aggressive (but ultimately unthreatening) NPC. A punch would have done the job, but in an attempt to role-play an evil character, I drew my sword and slashed the NPC, intending to kill him, which I assumed would be considered an evil action. The game responded as if I had punched the NPC, and I received "good" alignment points as a reward. The morality of violence is handled much as it is in other adventure games, in which the vast majority of violence (preemptive and otherwise) can be in some way interpreted as self-defense: attacking or killing unarmed civilians is marked as evil, but any entity designated by the game as an enemy is fair game, in the sense that no moral penalty is assigned for killing them. Some monsters are value neutral (giving no alignment points in either direction) while others, such as the werewolf-like "balverines" and generic "bandits," reward the player with "good" alignment points for every kill. The balverines, being replicable supernatural entities that cannot reasonably be dealt with through non-lethal means, are generally consistent with how such creatures are "morally" dealt with in the genre of horror fiction, but the bandits present a more problematic case. The bandits are human criminals, and Albion does not seem to have a functioning police force or justice system outside certain settlements, but it nonetheless gives me pause that their assassination is unquestioningly acknowledged as a universal

good—even if the bandit in question has not yet noticed the protagonist, and may not have yet attacked anyone else, the player is rewarded for firing an arrow into his throat from a great distance.

Similarly, it is common enough for adventure games to litter the landscape with treasures to be picked up by players. Some games apply this concept uncritically—many *Zelda* games allow players to wander into NPC houses, take money and smash objects, with no apparent response from the houses' owners—while others apply different rules to items found while "exploring" wilderness and items found while "exploring" private residences. *Fable* is at times very vague with the distinction, and since "examine" and "take" use the same key, I have often found myself "stealing" items by accident. At moments like these, the rules of both Albion and *Fable* itself can seem alarmingly random, and this randomness interferes with player experience by frustrating both the ability to grasp the intricacies of the rule system and the ability to maintain suspension of disbelief and become emotionally involved in the narrative.

Fable's authorial narrative is standard fantasy fare. The emergent narrative is slightly more ambitious, as various NPCs respond to the avatar's appearance (which reflects the protagonist's "attractiveness" and "scariness," as well as a visual index of the his alignment) and behave accordingly. However, the two rarely feel like elements of a cohesive story. The moral engine is what makes *Fable* an interesting game and a valuable object of study. Though the designers' stated aims were not realized, *Fable* can nonetheless serve as an effective blueprint for some of the more basic principles necessary to craft videogames that argue and critique elements of moral philosophy: specifically, a world allowing a high degree of freedom in which the actions taken by the

player are perceived, by NPCs in the gameworld and players outside it, to be morally meaningful. On this basis, the *Fable* engine is well-suited to the task at hand.

The philosophy espoused by Immanuel Kant holds that there is an *a priori* moral law that is inherent in consciousness itself, and can be discovered through reason alone. The fundamental principle of Kant's moral law is expressed in the categorical imperative. The categorical imperative—imperative because it is a command that a free being can choose to violate, categorical because it is not a means to any other end, but rather an end unto itself—is one principle with several formulations, the first of which is "act only in accordance with that maxim through which you can at the same time will that it become a universal law" (Kant 73). The morality of a given action is determined entirely in terms of its accordance with the moral law. This law cannot be deduced through empirical means, but only reached analytically through logic (Kant 77). Conversely, the morality of a given action is fundamentally unconnected to its empirical consequences, such as its ability to produce pleasure, or its accordance with the dictates of a religious order. Though these differing perspectives often reach similar moral conclusions about given actions, they are fundamentally incompatible with Kant's understanding of freedom.

Kant depicts human beings as rational animals, beings that deal with both the sensible (that relating to sensory perception) and intelligible (that relating to rational thinking). That humans either have free will or *believe* they have free will and act accordingly demonstrates the dominance of "practical" reason, that reason associated with action, over sensory "inclinations."[8] Inclinations, belonging to the world of the

8 Kant uses both of these terms repeatedly in several of his works, most notably in *Metaphysics of Morals*.

sensible, include not only physical stimuli such as hunger and pain, but also subjective emotional states such as happiness and "moral feeling." Because inclinations are subjective and ultimately subservient to the will, morality cannot be a matter of feeling. Rather, morality must be ascertained through rational processes, and the moral law must hold for all rational beings. Because it is logical, the law must not conflict with itself. The sanctity of free will is therefore the basis for Kantian moral law: autonomy must be protected in general, and can only be restricted when it restricts another being's autonomy. For Kant, this can be the only basis for morality: wrong actions are not wrong because they bring about unhappiness, or because they violate the will of God, or because they fail to conform to principles of moderation. They are wrong only because they are wrong. Moreover, actions cannot be said to be good merely because they *conform* to the law, because the actions could just as easily have been performed for another reason and conformed to the law only by coincidence (Kant 52-53). Rather, a moral action is one committed in conformity to the law, for no reason but respect for the law itself.

One particularly famous, and controversial, example of the categorical imperative is the prohibition of lying. The justification for this rule follows from the categorical imperative, in the sense that if everyone were permitted to lie, lying would cease to be possible. For Kant, this is problematic not for the consequences it would wreak (general erosion of trust), but because it represents a contradiction in terms. To lie, one must first have an assumption that one will be believed; lying, by definition, assumes a prohibition against lying. Illustrating the point with a hypothetical, Kant describes a man who needs to borrow money, but knows that he will me unable to repay the loan within the allotted time. Clearly, by Kant's standards, it would be immoral to do so, because a maxim such

as "when I believe myself to be in need of money I shall borrow money and promise to repay it, even though I know that this will never happen" cannot become a universal law without creating a contradiction:

> For, the universality of a law that everyone, when he believes himself in need, could promise whatever he pleases with the intention of not keeping it would make the promise and the end one might have in itself impossible, since no one would believe what was promised him but would laugh at all expressions as vain pretenses. (74)

In response to critics, Kant steadfastly defended his prohibition on lying even in the most extreme circumstances. *On a Supposed Right to Lie from Philanthropy* discusses another hypothetical case, in which a man is hiding in a friend's house from a murderer. The question at hand is, if the murderer comes to the door and asks if the man he is pursuing is hiding in the house, is the house's owner obligated to tell the truth?

It seems obvious, at first glance, that it would not be immoral to lie to a murderer in order save a human life. Kant, focusing on duty itself and not the immediate empirical consequences of following it, disagrees:

> Truthfulness in statements that one cannot avoid is a human being's duty to everyone, however great the disadvantage to him or to another that may result from it; and although I indeed do no wrong to him who unjustly compels me to make the statement if I falsify it, I nevertheless do wrong in the most essential part of duty in general by such falsification, which can therefore be called a lie [...] That is, I bring it about, as far as I can, that statements (declarations) in general are not believed, and so too that all rights which are based on contracts come to nothing and lose their force; and this is a wrong inflicted upon humanity generally. (612)

Lying to the murderer may protect one man from harm, argues Kant, but it would also directly harm every rational being in existence for its conflict with the moral law. Kant goes even further, suggesting that one cannot be certain that saying "yes" to the murderer's query would result in the friend's death, since the friend could have escaped

during the conversation, resulting in the murderer looking for him in the wrong place, and possibly being captured by neighbors in the process (613). Following the law does tend to lead to good consequences, but only if the law is followed out of duty: the law cannot be "hacked," to use some gaming parlance, for our convenience.

That any action could be moral *because* of its context contradicts Kant's ideal "kingdom of ends" in which every will acts as sovereign over the actions of every being:

> The concept of every rational being as one who must regard himself as giving universal law through all the maxims of his will, so as to appraise himself and his actions from this point of view, leads to a very fruitful concept dependent upon it, namely that of a kingdom of ends. [...] For, all rational beings stand under the law that each of them is to treat himself and all others never merely as means but always at the same time as ends in themselves. But from this there arises a systematic union of rational beings through common objective laws, that is, a kingdom, which can be called a kingdom of ends (admittedly only an ideal) because what these laws have as their purpose is just the relation of beings to one another as ends and means. (83)

In summary, Kant proposes a moral system derived from logic, based on the principle that what is right for one must be right for all. Such a morality is often counterintuitive, as in the example above, and it certainly bears no resemblance to the diegetic factors that motivate players and protagonists in most videogames. Were *Fable* to be reimagined based on Kantian principles, some effort would have to be made to teach the player what constituted morality in this universe. Kant, after a fashion, offers his own opinion on how this could be best accomplished.

In *The Metaphysics of Morals,* Kant lays out a plan for instilling his moral philosophy in youth. Differentiating his approach from that of the "dogmatic" method, in which the teacher alone may speak, and Socratic dialogue, Kant suggests a "moral catechism," in which the teacher proposes questions and the student is obliged to answer

him or her:

> The formal principle of such instruction does not [...] permit Socratic dialogue as the way of teaching for this purpose, since the pupil has no idea what questions to ask; and so the teacher alone does the questioning. But the answer which he methodically draws from the pupil's reason must be written down and preserved in definite words that can not easily be altered, and so be committed to the pupil's memory. (592)

Under this system, when a question is posed, a student must arrive at the answer through use of his or her own reason, and the teacher must arrange the questions to ensure that the student's reason does not go astray. When the correct answer is found, it must be committed to memory. This moral catechism bears a striking resemblance to the learning mechanism employed by most adventure games, which, despite being less linear than straight action games, nevertheless rarely give the player a legitimate, meaningful choice as to what choice of action to take. In such a game, the game poses a "question" to the player (in the form of a problem to be solved), and the player attempts to "answer," based on his or her reading of the game's rule system, aided by narrative's connection between the rule system and recognizable human experience. A correct "answer" allows the player to progress, and must often be repeated many times to ensure the player can reproduce the answer by memory when he or she is called upon to do so.

By Kant's standards, the existing *Fable* tutorial is deeply problematic. First, that the protagonist is paid for good deeds allows him to operate from an inclination toward greed, and act in accordance with the law without understanding why he ought to do so. Rather, the fulfillment of duty must be directly connected to the best win condition (there must, after all, be several), and it must be represented non-diegetically, i.e. it must have no empirical value *in the gameworld itself.* Though the two persons of the avatar, the

player and the protagonist, will both to some extent share their sensible perception of the world, only the player has the outright power to command. Therefore, elements outside the game narrative are a more effective representation of the intelligible world than elements inside it.[9] Actions must have empirical consequences as well, of course, and it will be important to develop a diegetic means to mark the happiness of NPCs.

Second, the father expects the child to intuitively know what is right and what is wrong—presumably through moral feeling—and gives little instruction on *how* to pursue the good. It's possible that he gave the protagonist more specific moral training when he was younger (i.e. before the game narrative begins), but it is not present in the text as is. Consequently, players face the often bizarre results discussed earlier. A Kantian perspective, introduced in the narrative and reinforced through gameplay, would be one way to provide consistency. In the case of the philandering husband, clearly the *worst* option, in terms of the categorical imperative, would be to accept the bribe and inform on the husband anyway, even though this may seem largely indistinguishable, in terms of consequences, from refusing the bribe in the first place. (The money earned by the player in the tutorial section of the game ceases to exist once the tutorial concludes.) The conflict involving the barrels must not allow the player to perform bad actions and yet receive a reward without an internally consistent narrative explanation as to why, simply to avoid stumbling into nonsense; beyond that, the primary moral option is obvious, as Kant's perspective clearly opposes the breaking of promises and the unlawful destruction or seizing of property. In the conflict involving the teddy bear, the issue of intent must

9 I must note here that Kant defines the intelligible as impossible to represent. I can only respond that *Fable* is a text, and texts represent things. There is nothing else they do. Nonetheless, I cannot deny that Kant would likely have found this idea appalling.

first be resolved before a clear perspective can be discerned.

As it is currently written, this scenario presents the player with three options: hit the bully and give the teddy bear to the little girl, hit the bully's victim and give the teddy bear to the bully, or hit the victim and give the bear to the little girl. The first is obviously the "good" solution, and the second is obviously less desirable, from a moral perspective. The third is harder to interpret, since it makes little sense. What kind of intent would logically lead to the performance of those actions? One possible explanation is that the protagonist (more accurately, the player) changed his or her mind; if this is the case, this will need to be acknowledged by the narrative, and the player will have to make a conscious decision to apologize or otherwise acknowledge that his or her intent has changed. Giving the protagonist the ability to speak, and the player to manipulate this speech through some method (such as a dialogue tree), would be useful for this purpose. At any rate, using Kant as a model, only the best solution would qualify as a good deed, while the others would be failures of varying severity, a trope that would need to be consistent across any conflicts designed to have multiple parts. In addition, failure to conform to the moral law must have consequences (primarily non-diegetic, but possibly diegetic as well) to the player that are greater than the consequences resulting from the actions' resulting pleasure or pain.

Pleasure and pain will need to have a direct effect on the game experience. Two ideas could be applied to this end. In the first, empirical consequences (i.e. pleasure and pain) would be predictable consequences of most actions; in the second, the consequences of a given action would be randomized. Each solution yields a slightly different argument.

If the empirical consequences of actions are predictable, situations could be designed in which strict conformity to the moral law will produce either pleasure or pain for the public (in the form of NPCs). In the former case, the *a priori* moral law happens to produce the best empirical results, and there is no conflict between the two. The latter case is far more interesting, since it places the two in direct conflict. Kant argues that this scenario is the test of a good will: the ability to obey the moral law even when one is strongly tempted, by inclinations such as moral feeling or popular sentiment, to do otherwise (591). On a more general level, arguing for a moral perspective that goes beyond mere expedience will necessitate allowing the protagonist to be punished for doing the right thing.

If the empirical consequences of actions are randomized, on the other hand, the player will have a radically different experience that is Kantian on a more fundamental level. In this arrangement, the happiness of others is acknowledged as a good—Kant connects it with the struggle for self-perfection as part of the moral law (517)—but also holds that happiness, belonging to the world of the sensible, cannot be predicted *a priori,* and thus cannot be the *basis* for moral law. A player of this game would quickly learn that honors bestowed on the protagonist by the others for increasing their happiness, while enjoyable to the protagonist and useful to the player, must not be confused with morality itself. Whether empirical consequences are predictable or randomized, the gameplay conventions of *Fable* present one major problematic issue: the morality of violence.

Kant, frames human relations in terms of a lawless state of nature on one hand and government, presumably one resembling the governments of 18th century Europe, on the

other. Property right is a concept that properly belongs only to government, and it can be assumed that prohibitions on violence are also absent from a state of nature. (*Fable* does allow players to buy property, but its actual use value is minimal, and as such I do not intend to dwell extensively on property right.) It could be argued that an agreed-upon prohibition against certain types of violence is, in fact, the *definition* of society, and that that agreement itself removes humans from the anarchic state of nature. Albion is clearly not a state of nature, but it also does not resemble a state under a government as described by Kant. There are civil laws, certainly, and the idea of the "criminal" exists, but enforcement of these laws does not extend to large sections of the game map, and while police can (and do) harass, fine and/or attack the protagonist for wrongs he has done to others, he cannot appeal to any external authorities for the same privilege if he is wronged by a third party.

Moreover, it is unclear how a concept like self-defense fits into the categorical imperative. Kant writes extensively about the importance of capital punishment for murderers, but offers little about whether the moral law allows the lawful killing of those *attempting* murder. Because self-defense is the basis for most of *Fable*'s violence—and therefore most of the game in general—I am positing a right to defensive violence that I believe to be consistent with the categorical imperative: a person who has not committed a crime may resort to lethal violence to prevent murder, and the party attempting murder forfeits his possessions upon his death. The extent to which pre-emptive attacks are thus justified on characters marked as "criminals" that have not been shown to have committed a specific crime warranting capital punishment varies on whether it can be known, *a priori,* that a group of armed NPCs intends to do the protagonist harm. This is

a particularly important issue because, under the rules set forth by Kant, positive consequences resulting from a dead "undesirable" are not sufficient to justify the act of killing.

Every action, including every kill, must be consistent with the moral law, and the consequences of actions that contradict the law must be far less reversible than the empirical consequences of such actions. *Fable* allows the protagonist's reputation to follow him, to some degree, but this reputation is based only on the total sum of the player's actions. The damage to a player's alignment, and therefore reputation, through evil actions can be undone by performing good actions. Assuming the NPCs care primarily about empirical consequences (i.e. the happiness of their community), this is a plausible model, but the moral law is not so fickle. The game's reward system must ultimately be tied into a non-diegetic representation of morality that does not allow players to "buy back" evil actions with good ones. The pain caused by an evil action can be counteracted, but damage to humanity itself cannot.

Adherence to the moral law, then, must be represented, because videogames necessarily involve representation. It must also be represented in a non-diegetic fashion, so it cannot be interpreted as "empirical" within the world of the protagonist. The player's "alignment," measuring adherence to the law, need not be accessible to players at all times, though there are certainly advantages to doing so. The only time when it is necessary to represent the law to the player is after the game's conclusion, and the protagonist's "conclusion" as well: the protagonist's afterlife.

Just as consciousness necessarily leads to the moral law, Kant claims that the existence of the moral law necessarily suggests the idea of a perfect being, a God (240).

Similarly, that adherence to the law is imperative yet ultimately impossible to fulfill suggests the existence of an afterlife of perpetual improvement (238). The gameplay possibilities suggested by the necessity of an afterlife are myriad, but beyond the scope of this discussion, if only because Kant is understandably unclear on what such an afterlife would entail. It is sufficient to note that an afterlife must exist at a narrative level, whether or not any gameplay is involved, and that this afterlife is the point where the non-diegetic, non-empirical consequences of the protagonist's actions become diegetic, and therefore visible to the protagonist himself.

The existence of a God or gods and an afterlife, often treated with empirical, scientific certainty in adventure games, is not strictly necessary in a utilitarian model. In fact, utilitarianism does not rely upon a knowledge of the universe itself attained through reason, as does Kantianism. Utilitarianism, by attempting to derive morality from empirical knowledge and human sensibility, offers a very different perspective, which in turn leads to very different design rules.

While Kantian philosophy owes much to Greek and Christian conceptions of logic and morality, it can be adequately described in terms of the writings of the individual for whom it is named. Utilitarianism, named for a property of the theory itself and not one of its proponents, has a wider canon. In this chapter, I will deal with two primary texts: Jeremy Bentham's *An Introduction to the Principles of Morals and Legislation,* and John Stuart Mill's *Utilitarianism.* The utilitarianism espoused by Bentham and Kant is preceded by Greek theories of hedonism, and would later be divided into more specific variants such as act-utilitarianism and rule-utilitarianism. However, I will be dealing primarily with the texts detailed above, which (despite their differences) I believe are

sufficiently integrated to form the basis of a single videogame.

"Nature has placed mankind under the governance of two sovereign masters," writes Bentham: "*pain* and *pleasure*" (1). In this sentence lies an exceptionally concise summary of utilitarianism. Pain and pleasure, far from animalistic inclinations to be subordinated by the will, are depicted as dominant. What Kant subjugated to the realm of the sensible, the utilitarian interprets to be the most fundamental fact of existence, and the source of all morality. Expanding on Bentham's work, John Stuart Mill suggests that any moral philosophy must necessarily begin with observation of that which is self-evident, rather than deduced abstract principles:

> The truths which are ultimately accepted as the first principles of a science, are really the last truths of metaphysical analysis, practised on the elementary notions with which the science is conversant; and their relation to the science is not that of foundations to an edifice, but of roots to a tree, which may perform their office equally well though they be never dug down to and exposed to light. [...] When we engage in a pursuit, a clear and precise conception of what we are pursuing would seem to be the first thing we need, instead of the last we are to look forward to. A test of right and wrong must be the means, one would think, of ascertaining what is right or wrong, and not a consequence of having already ascertained it. (2)

By assuming the primacy of what can be known through observation, the principle of utility—"that principle which approves or disapproves of every action whatsoever, according to the tendency which it appears to have to augment the happiness of the party whose interest is in question"—is the only logical place to start in defining a moral philosophy (Bentham 2). Bentham defines the terms quantitatively, measuring pleasure and pain in terms of "intensity," "duration," "certainty or uncertainty," and "propinquity or remoteness," (29) while Mill adds a qualitative distinction, suggesting that intellectual pleasures are inherently superior, and more valuable, than their bodily

counterparts (Mill 7).

Put simply, a utilitarian *Fable* will necessitate a moral engine that quantifies the pleasure and pain that result from actions and determines their morality accordingly. As in the Kantian model, the narrative will have to be sufficiently integrated with the game's ethics to make actions appear morally significant. However, this will require less focused explanation in the utilitarian model, since the happiness of the world's inhabitants should be readily apparent to the player, and increasing the happiness of NPCs naturally lends itself to diegetic rewards. Many games rely on such incentives: an NPC, pleased by a player's previous action, rewards the protagonist, allowing the game to progress. Most adventure games, after all, conclude with the hero having made the world a demonstrably happier place. The designers may "piggyback" on existing genre conventions, needing only to take care to note that the ideal is global utility, not personal gain. Failure to treat the happiness of others as a goal in itself, as opposed to a means to other goals, could plunge the game's moral system into egoism. If aggregate happiness is a tangible good, it must be treated as the game's goal.

Once this is established, however—possibly by directly referring to Bentham's rules, upon which the moral engine could be easily coded—very little in the way of tutorial play would be necessary. Unlike Kant's chatechistic model, which mirrors the linear building of complexity found in most adventure games, utilitarianism suggests an approach more reminiscent of sandbox games: set the rules in motion and let the player experiment. If the player finds a counter-intuitive solution that the designers did not foresee, all the better.

For this reason, the *Fable* tutorial, which was made more restrictive for the

Kantian model, becomes less so for a utilitarian model. The three tutorial scenarios described earlier—the stolen teddy bear, the philandering husband, and the merchant's goods—are much less intuitive from a utilitarian standpoint. The stolen teddy bear presents two decisions: how to acquire the bear, and what to do with it afterwards. While the "best" solution by utilitarian standards would likely be the same as the "best" solution in the existing *Fable* (hitting the bully and returning the teddy bear to the little girl), complications might nonetheless arise. While the little girl seems to have the most right to the bear, what if it could be known that the bully would appreciate it more, or that his continued possession of the bear would create more happiness than returning it to the little girl? What are the consequences of hitting the bully, as opposed to the little boy? If, even after the bully has been run off, the little boy refuses to hand over the teddy bear, the player would have no choice but to attack the little boy anyway. Would attacking the little boy *first,* thus committing an act of violence against only one person, be preferable than attacking them both?

Similarly, the philandering husband requires a more complex reading if aggregate happiness is to be the goal. The wife clearly suspects her husband, and seems to be bothered more by his continued absence (and consequent inability to help with the household) than by his infidelity. However, it could still be argued that the wife would be happier not knowing the truth about her husband's activities, and the husband would certainly prefer this outcome. He prefers it so much, in fact, that he's willing to pay for it. Under these assumptions, taking the bribe and staying quiet could be seen as the most moral solution. If we assume that the wife would, in fact, prefer to know, the issue is still not resolved. If the player chooses to accept the bribe and inform the man's wife anyway

—the *worst* option from a Kantian standpoint—it might lead to greater overall happiness than refusing the bribe. After all, the amount of money involved is a pittance to an adult, but crucially important to the child protagonist, and doesn't a man who attempts to bribe a child to deceive his wife deserve to lose the money he offers? An option to accept the bribe, tell the truth, and then return the bribe to the wife (who will, presumably, use it more responsibly) might produce an even better result.

It should be noted that the iterative nature of videogames would make it easy for players to replay the same scenarios repeatedly, looking for the best results. Designers of a utilitarian *Fable* would be wise to allow frequent saving and loading of game-states to exploit this feature of the medium. Counter-intuitive solutions, such as taking the philanderer's money and squealing anyway, will produce unique results, and these results may be compared through repeated play. Unlike in one version of the Kantian model, however, the consequences of any action must be fixed—players must have a reasonable ability to predict the consequences of their actions. Without this ability, utilitarianism becomes untenable as a philosophy, whether on Earth or in Albion.

Despite this restriction, the utilitarian model offers the possibility of a causality better suited to long-term plans. After all, unlike Kantian morality, utilitarianism has no particular problem with treating people as means to an end rather than as ends in themselves, as long as the net gain in happiness is sufficient. That an individual action causes pain does not necessarily make it a bad choice, as certain actions will produce different results depending on the larger series of actions of which they are a part.

Responding to claims that it is not feasible to expect people to consider the happiness of society as a whole in their actions, Mill writes:

> [T]he motive has nothing to do with the morality of the action, though much with the worth of the agent. He who saves a fellow creature from drowning does what is morally right, whether his motive be duty, or the hope of being paid for his trouble; he who betrays the friend that trusts him, is guilty of a crime, even if his object be to serve another friend to whom he is under greater obligations. (17)

This produces an objection: what if someone saves a drowning man for the purposes of slowly torturing him to death? What if, on the other hand, someone betrays a friend's trust because lethal consequences would result were that trust maintained? Mill answers:

> I submit, that he who saves another from drowning in order to kill him by torture afterwards, does not differ only in motive from him who does the same thing from duty or benevolence; the act itself is different. The rescue of the man is, in the case supposed, only the necessary first step of an act far more atrocious than leaving him to drown would have been. (61)

Mill goes on to clarify the distinction between intention and motive, defining intention as "what the agent *wills to do*" and motive as "the feeling which makes him will so to do" (61). Using Mill's terminology, motive would be irrelevant in a utilitarian *Fable,* and thus the game engine would not need to know such information in order to respond to it. As for intention, it might be necessary to force players to establish their intentions before a given action, but it would likely be sufficient to "deduce" player action from the actions that follow it, as in the example of the drowning man. Since the player, unlike the avatar (or a real person), could be privy to her overall contribution to the happiness of the human species at any given time, she need not wait for a chain of actions (e.g. saving a drowning man, then torturing him to death for some reason) to be completed to see that the end result of a chain of actions is the sum total of the results of the actions within it—and that a positive action might retroactively become a negative

action, or vice versa, depending on the actions that follow it. While this might allow players to perform good acts without good motives, the engine is designed to track the morality of actions, not the moral worth of the actor. Mill's "motive," in the case of a videogame, calls to mind the dual nature of the avatar. The protagonist often has motives that the player does not share. The "worth of the agent," to use Mill's phrase, depends on a factor about which the game itself neither "knows" nor "cares": the state of mind of the player. When the player performs actions that lead to increased happiness without intending to do so, or without communal happiness being his primary aim, she is behaving morally, but only when she performs actions that promote happiness *because they will promote happiness* has she demonstrated her worth as a moral agent. In the Kantian *Fable,* scenarios must be carefully designed to devalue actions taken in conformity to the moral law, but not out of respect for the law itself, but to make such a distinction in the utilitarian *Fable* would, in and of itself, contradict Mill's division between motive and intent.

There are several such instances, in which a design element easily amenable to one moral philosophy becomes more complicated in another, or an element crucial to one becomes irrelevant or necessary to exclude in another. Growing as they do from such radically different principles—duty to an *a priori* moral law for its own sake, and aggregate human happiness—it is unsurprising that Kantianism and utilitarianism yield different answers to a great many moral questions. This is not to suggest that there are no agreements to be found between the two philosophies, however. The study of morality tends to focus on the points of conflict, because nearly any moral philosophy will yield similar answers to the normal, day-to-day experiences that make up the average human

life. It is likely that, given a sufficiently mundane design (at both narrative and ludic levels), adaptations of *Fable* based on Kantian or utilitarian principles would appear to have only superficial differences. To create such games, or even to design them at a conceptual level, would be something of a waste of time. To make the philosophical basis of the games' morality engines most meaningful, designers must ensure that players are routinely forced into situations in which the world's morality at first appears counterintuitive. These situations are most apparent when two competing philosophies are placed in contrast to one another. Hence, I will now compare the Kantian and utilitarian *Fable*s directly, focusing on two key issues with which each game must necessarily deal: lying, and justice.

Standing apart from theft and murder, lying is a generally agreed-upon sin that is largely absent from videogames, for the simple reason that lying requires the ability to converse. Most games feature dialogue of some sort, but the player's freedom of action is invariably limited by the technical and creative limits of modern videogame design. Either the conversations are scripted outright, and the player becomes an observer, or the player is allowed to choose between several scripted options (a dialogue tree), which often have only superficially different results.[10] In order to lie, conversation itself must first be an option.

I do not believe, however, that the current limitations on videogame conversation make morally meaningful arguments about lying impossible. The dialogue tree, limited as it is to games that allow both conversation and limited (but meaningful) control over

10 As always with such sweeping statements, there no doubt exist some counter-examples, but these are the most common models for videogame dialogue in single-player games.

player dialogue, allows players to lie in many games. However, even when morality is specifically part of the game's rule system, lying is not always treated as being morally significant. I recall a frustrating experience playing Troika's *Arcanum* as a character I intended to be "good," and finding it impossible to complete a given task without lying, knowing that the game would not consider my character less good for such an act, and also that my character would receive no credit for failing to the achieve the goal out of a refusal to lie. This, therefore, must be addressed in either *Fable* adaptation: the distinction between truth and falsehood must be a morally meaningful one.

As discussed in *On a Right to Lie from Philanthropy,* the categorical imperative excludes lying as a morally permissible option even under the most dire circumstances. Lying to achieve a goal that is, in itself, laudable must negatively affect an avatar's morality in the Kantian *Fable.* The consequences of such a lie, for reasons discussed earlier, can indeed be quite positive, or quite negative, and may manifest in any number of ways. There is no reason that all NPCs inside the text need subscribe to the moral arguments put forth by the text, after all, and thus the protagonist might not even be considered untrustworthy as a result of certain lies. The non-diegetic consequences, of course, must remain. Conversely, the utilitarian perspective considers a lie that increases the world's total happiness to be a categorical good, and the consequences *must* be made visible empirically—though the outcome of such a lie might be difficult to predict.

The most direct way to illustrate these principles is that supplied by one of Kant's critics quoted in *On a Supposed Right to Lie from Philanthropy*: a frightened man comes to a friend's house and asks to hide out from a murderer. The friend allows the man into

his house, and the murderer subsequently arrives, and asks if the man is inside.[11]

The correct response is obvious, and different, from both philosophical perspectives. The Kantians say yes; the utilitarians say no. However, the problem does not end with the player's answer.

What if, asks Kant, the frightened man has, unbeknownst to his friend, slipped out the back window during the conversation and begun to flee through a field? If the friend tells the murder what he believes to be the truth—that the marked man is, indeed, inside the house—the murderer will no doubt search the house, allowing the frightened man to get further and further away, and increasing the likelihood that the murderer will be apprehended before any more damage is done. If, on the other hand, the friend has "lied" to the murderer, by giving an answer contrary to what he *believed* to be true, he has then *increased* the likelihood that the frightened man will be killed (612). The fate of the man in question is not Kant's primary concern, of course, and he merely uses the example to point out that consequences in the sensible world cannot be predicted *a priori*. Additionally, he suggests that the friend (or, in our current focus, the player) is fundamentally complicit in the death of a man as a consequence of a lie, whereas he is ultimately innocent in the death of a man as a consequence of telling the truth (612-613). The Kantian answer is simple. It is radically contrary to most people's sense of morality, but it is simple, and it requires no particular skill to execute. The utilitarian answer, when consequences are made difficult to predict, is perhaps more satisfying to the modern mind, but it is also more difficult. The utilitarian model must textually allow the player

11 This hypothetical scenario does not allow the friend to respond with anything other than a yes or no answer. Some convincing reason for this restriction of action might be useful, but since players are rather used to restricted action in videogame conversations as it is, it might not be necessary.

to figure out, with some skillful reading of the game text, the correct action; the Kantian model need not. Lastly, it must be noted that conventions of the adventure genre would likely suggest to players that the correct course of action, whether or not the player tells the truth, is to kill the murderer on sight. This problem *will* need to be addressed, in a narratively convincing fashion. Kant, perhaps, was not describing a society as ubiquitously violent as that of the modern adventure game.

Another major point of conflict between the two perspectives is that of justice. H.J. McCloskey suggests a scenario in which a hypothetical utilitarian is visiting a community torn by racial strife (qtd. in Rachels 110-111). In this community, a black man rapes a white woman. Riots ensue, as white civilians swarm through the streets beating and killing blacks. The hypothetical utilitarian knows that he will be believed if he claims to have witnessed the crime, and that this false testimony will end the rioting. In this scenario, is the utilitarian morally obligated to falsely accuse someone in hopes of ending the riots and restoring public safety? James Rachels summarizes the conflict as follows:

> The argument is [...] that if someone were in this position, then on utilitarian grounds he should bear false witness against the innocent person. This might have some bad consequences—the innocent man might be executed—but there would be enough good consequences to outweigh them: The riots and lynchings would be stopped. The best consequences would be achieved by lying; therefore, according to Utilitarianism, lying is the thing to do. But, the argument continues, it would be wrong to bring about the execution of an innocent man. [...] Justice requires that we treat people fairly, according to their individual needs and merits. The innocent man has done nothing wrong; he did not commit the rape and so he does not deserve to be punished for it. Therefore, punishing him would be unjust. (111)

To be sure, it can be argued that utilitarianism need not support such an action, or

generally conflict with ideas of justice. John Stuart Mill, discussing the concept of justice in *Utilitarianism,* offers his own definition:

> It appears from what has been said, that justice is a name for certain moral requirements, which, regarded collectively, stand higher in the scale of social utility, and are therefore of more paramount obligation, than any others; though particular cases may occur in which some other social duty is so important, as to overrule any one of the general maxims of justice. (59)

However, this definition does not solve the problem detailed above, for it leaves unclear the question of whether this problem happens to be one of those particular cases. Others have made distinctions between ways of putting utilitarianism into practice by making a distinction between act-utilitarianism (in which every action is judged on its consequences) and rule-utilitarianism (in which actions are judged by their overall *tendency* toward consequences), but our analysis concerns only the former.

This point of conflict is, after a fashion, a natural outgrowth of the previous one: a challenge to the idea of the noble lie that is more plausible, and more disturbing, than the murderer at the door. The lie, in this case, is a larger one, told to the community at large instead of a particularly violent individual. From a Kantian perspective, any lie makes victims of the whole world, but assuming a more conservative view of lying and its victims, there seems to be a categorical difference between a lie that results in the frustration of a murder attempt and a lie that results in the death of an innocent. The Kantian perspective is identical for both: one must not lie, even if to do so will save lives. The utilitarian must either embrace the moral necessity of the death of an innocent, or find another reason to reject it on the basis of aggregate happiness.

Assuming, for the sake of argument, that this cannot be done, a scenario like this

one could be easily designed in an adventure game like *Fable*. However, it need not be presented as a simple binary, like the murderer at the door. Regardless of its morality, the solution to the rioting "problem" suggested in McCloskey's hypothetical is certainly a creative one. Players who opt for such a solution in the absence of direct prompting ought to be rewarded if this result is consistent with the moral rules of the game text: this might be seen as an example of the kind of emergent behavior that a utilitarian engine might encourage. But even then, there might be further long-term consequences that outweigh the short-term benefits to the community, and a utilitarian *Fable* would need to address this possibility. The Kantian model, of course, has but one solution, regardless of how the empirical consequences play out.

In the Kantian and utilitarian models I have proposed, the same design principles apply: delineate a "correct," or ethical, way to play, and reinforce these ethical rules through both gameplay feedback and narrative. The rules, if supported by the narrative, will appear plausible and meaningful, and as in a work of literature, the player will have the ability to make connections between the fictional (or virtual) world and the real one. This act of making connections is not an "effect" as the term is usually applied to videogames, but is rather more like that of a children's fable, or any story (for any intended audience) in which the author seeks to convey to the audience that a certain action was moral or immoral for a certain reason.[12] But these games need not be so straightforward: if games can make arguments directly, they can also do so satirically.

To accomplish this feat, the formula must simply be reversed. The player must be

12 This is rather similar to the "meanings" model proposed by Henry Jenkins in "The War Between Effects and Meanings," to be discussed in more detail in the conclusion.

compelled to play in a way that is explicitly defined as nonsensical or immoral by the narrative. This will make the rules appear arbitrary, and the protagonist will appear foolish for following them. If the Kantian model were to reward the player non-diegetically for displaying adherence to the moral law out of duty, while narratively emphasizing only the pain such behavior causes and repudiating the afterlife Kant claims is suggested by the law's existence, players would leave behind a ruined world that would have been far happier had they never entered it. Players would be assured of the rightness of their (intuitively appalling) actions only by the fact that they would have received a good moral "score," a score the narrative had depicted as meaningless. Conversely, if the utilitarian model were to emphasize Kant's moral law at the narrative level, players would nonetheless be forced to break it constantly to satisfy the game's ethics and advance—only to be reassured of the meaninglessness of the happiness they sowed when all the world's inhabitants find their souls surviving death, and the empirical consequences of their actions rendered irrelevant.

Anyone who has spent a sufficient amount of time reading moral philosophy will eventually come to the conclusion that, regardless of the value such debates may have, life tends not to present us with such clearly delineated moral dilemmas. In the next chapter, I explore moral issues raised by the recent "War on Terror," suggesting ways in which videogames may argue moral viewpoints in a way that are less academic and more immediately relevant to the public at large.

Chapter 3: America's "War on Terror"

The games described in the previous chapter deal with abstract philosophical principles. While the moral issues raised are specific to the gameworld itself, they are quite general in terms of our own world, essentially functioning as a space to "play out" hypothetical scenarios. Kantianism and utilitarianism differ dramatically on a wide variety of issues, and any videogame that attempts to model either will necessarily need a fairly wide range of available player actions. A game like *Fable,* taking place as it does in a world that is similar to our own, but recognizably different—if the inconsistent and uninspiring moral law does not dispute the latter claim, the presence of werewolves and giant arachnids should—is useful for the studying of moral principles at a distance.

However, many games rely on tropes and settings derived from external sources to form their narratives, and reality is one of those external sources. Gee notes that, in the wake of the terrorist attacks of September 11, 2001, numerous commercial and non-commercial videogames dealing with war in the Middle East were released (148). Works of fiction such as action movies are often inspired, however loosely, by current events. America's "War on Terror" might have gone so far as to produce narratives that have, in turn, colored the public's perception of the conflict itself: it has been argued that Americans' debate over the morality of torture has been influenced by the television show *24.*[13]

Many of the post-9/11 games to which Gee referred are simple Flash-based games with no function beyond simple, arcade-style gameplay and the spectacle of violence itself—the political equivalent of "peeing Calvin" bumper stickers. Others

13 Jane Mayer explored this idea on several levels in "Whatever It Takes."

follow the genre conventions of a genre that traditionally has treated narrative in a superficial manner, and avoid dealing with moral issues at all by centering the action on conventional warfare (as opposed to the more complicated realities of military occupation) without regard to real-world context.

If the principles I've espoused hold, and the games I've described in the previous chapter can model and critique moral philosophies, then a slightly more realistic narrative and specifically tailored rule system could yield games that offer specific perspectives on current issues, as well as the means to critique those perspectives. This chapter focuses on America's "War on Terror," and two-specific issues that have arisen from it: the blurring line between civilians and soldiers, and morality of torture as a means to interrogation.

The games I will describe in this chapter would be considerably narrower in focus than the hypothetical *Fable* mods described earlier. Kantianism and utilitarianism are useful examples largely because they are so contradictory, and as a philosophical exercise, it makes sense to view them in isolation. In the games I will describe in this chapter, designers need not refer to an existing school of moral philosophy, and can freely integrate ideas from multiple traditions, to the extent that this is possible while maintaining internal coherence. Put more succinctly, the previous chapter discusses ways to materialize existing philosophies; this chapter concerns the materialization, and subsequent critique, of new ones.

Few games, thus far, have dealt with the civilian/soldier problem or torture at all, let alone very effectively. In this chapter, I will deal primarily with two videogame texts: *Command & Conquer: Generals,* a real-time strategy (RTS) game set in the modern

military era, and *The Punisher,* an action game based on the Marvel Comics franchise that deals explicitly with torture.

Civilians have long been featured in war-themed action games. Generally, avoiding civilian casualties is designated as an ethical goal, and one that can be achieved without great difficulty. One notable exception to the former is *September 12th,* a free Flash game in which players attempt to kill terrorists while avoiding civilian casualties. The problem—indeed, the central argument of a game explicitly designed to *make* an argument—is that it is impossible to do both of these things at the same time. Whenever a civilian is killed, more terrorists are added to the map. *September 12th* is, self-consciously, a game that cannot be won.

The treatment of civilians in *Command & Conquer: Generals,* a more conventional, commercial game, is not possessed of the same subtlety or pique. One possible reason is that the armies the player controls do not make an entirely clear distinction between civilians and soldiers. The game offers play from three perspectives: the United States of America, China, and the "Global Liberation Army" (GLA), a generic middle-eastern (implicitly Islamic) terrorist network.

In *Generals,* the US army is represented in game's iconography by the color blue. The US has the most technologically advanced equipment (including a healthy dose of science fiction) and the best training for individual soldiers: units[14] are expensive, slow to produce, and more specialized than those of opposing armies. Each unit type has audio cues that play when selected, a snippet of speech from the unit (or pilot) to whom the player wishes to assign orders. The US units' audio cues, delivered in the style of John

14 "Units," in this case, refers to all mobile forces, infantry and vehicles.

Wayne, are ham-handedly heroic and patriotic: "doin' what's right," "protecting democracy," etc.

China, represented by the color red, offers a different approach. Units are faster and less expensive to produce, allowing the Chinese army to overwhelm opponents with sheer numbers: rifle infantry, for example, can be produced either two or eight at a time. Certain units become more effective in large groups, and this effectiveness can be further bolstered by investing in the "nationalism" tech upgrade. The Chinese army also possesses the unique ability to heal injured units by means of propaganda towers, which in turn can be made more effective through the use of subliminal messaging. In general, the Chinese army relies on conventional force using 20th century technologies, bolstered by ideology and group identity, as well as more controversial cold war-era weapons such as napalm, land mines and nuclear missiles.[15] Chinese units' audio cues emphasize national identity: "defending China," "we are the Red Guard," etc.

The GLA, represented by the color green, relies on stealth, versatility, and unconventional attacks such as hijackings and suicide bombings. Many of the GLA's units are explicitly designed to exploit opponents' units: destroyed vehicles can be salvaged for GLA upgrades or sold for scrap, and the GLA can receive monetary rewards for every kill made in the cause's name. The technological needs of the GLA are lower than either opponent, as GLA structures require no electricity to operate, and are assembled by human workers instead of machines. These workers mark a departure from the standard *Command & Conquer* formula, as they are clearly civilian non-combatants.

15 In the interest of game balance, the nuclear weapons are astonishingly weak, but realism is clearly a secondary concern, if that.

China and the United States are both nation-states with conventional armies (and consequently all their units are military personnel), but the GLA, lacking a single state from which to draw an army, shows considerably more variance in its ranks. None of the units have particular state alliances, and the game manual describes the rifle infantry unit (the "Rebel") as having "more spirit than training or equipment" (28). Other units are designed to blend in with civilian non-combatants in urban areas. Two units show particular significance: the "worker" and the "angry mob."

The "worker" is a builder and resource gatherer. Worker units are cheap to produce, but go about their jobs slowly, gathering supplies to be converted into cash, repairing damaged structures, and building new ones. The U.S. and China split these duties among two unit types each, and both are high-priority targets when invading an enemy base. Combining the two functions makes the workers absolutely crucial to the GLA's war effort, and makes their elimination a highly effective tactic for opponents. However, the workers themselves have no combat abilities whatsoever, and unlike the opposing armies' professionals operating machinery for building and gathering, do not seem to offer their services on an entirely voluntary basis. The workers' audio cues imply both extreme poverty and coercion: "don't hurt me," "I need shoes," etc.[16]

The "angry mob" unit, while explicitly voluntary in nature, also presents an unusual case: the angry mob is a crowd of civilians armed with guns and homemade explosives. Though comprised of many polygonal "people," the angry mob acts as a single unit, and becomes less effective as individual members are killed off. However,

16 An expansion pack, entitled *Command & Conquer: Generals: Zero Hour,* allows the player to purchase shoes for the workers, improving their movement speed. While this is tactically useful, the use of shoes to accomplish this bonus is likely a tongue-in-cheek reference to the audio cue in the original game.

civilian non-combatants can also join the mob to replenish their numbers.

On a functional level, the worker and angry mob are no different than any of the game's other units. They are produced/recruited in the same manner as any other unit. However, *Generals* is not an exercise in abstract symbolism. Certain recognizable symbols are intentionally mapped to certain gameplay elements. The communist Chinese are red, while the GLA's green calls to mind the Ottoman empire, and (in popular consciousness) Islam in general. The technological and tactical differences between the three armies carries within it a kind of moral hierarchy: as powerful as the space-age American weapons are, they are not nuclear, biological or chemical in nature. The Chinese use of propaganda, napalm, land mines and nuclear weapons (complete with unrealistically short-term radioactive contamination) calls to mind many morally questionable elements of the cold war era. The GLA's arsenal of chemical weapons, hidden proximity explosives, car bombs, hijackers and suicide bombers call to mind tactics frequently reviled as cowardly in current "War on Terror" rhetoric.

This should not be interpreted to mean that this moral and technological hierarchy is the only way to read the game. Beyond the events depicted in the battles, there is no objective viewpoint and little verbal narration common to all three perspectives. Like Gee's claim about Shadow in *Sonic Adventure Battle 2,* any army commanded by the player is, provisionally, righteous. During gameplay, players hear only from their own troops and occasionally from disembodied advisors. The opposing troops are silent at all times, and the fog of war (a common genre convention that limits players' ability to observe enemy movement) ensures that most attacks, however inevitable, will come as a surprise. When an opponent, human or computer, eviscerates one's carefully planned

defenses with an unexpected tactic, it is easy to feel (as a player) that the opponent is "cheating," and the technological differences reinforce this idea at a narrative level. Just as one thinks (rather irrationally) that the rules of the game have been broken by the opposing player, Western cultural conceptions of warfare reinforce the idea that the rules of war have also been broken, this time by the opponent's army. A player used to the earlier *Command & Conquer* games will be surprised and alarmed by the ease with which a large (and very expensive) fleet of tanks can be reduced to heaps of smoking metal by a small group of inexpensive, low-tech GLA suicide bombers. It is a small leap, assuming the player is responding to the game on both narrative and ludic levels, to thinking of the GLA forces are cowardly and immoral. The feeling of being constantly under assault by a cruel and unforgiving force is a crucial part of the RTS formula, and the political identities of the parties involved help play into that feeling.[17] However, because players can experience battles from multiple perspectives, controlling any of the three armies against any opponent or opponents, these feelings are ultimately undermined: a player controlling the GLA will have quite different ideas about the rules of war when the U.S. stages a blitzkrieg using vastly superior technology.

Ultimately, the balance required to make *Generals* function well as an RTS game cannot help but place the three armies on roughly equal footing, in terms of both morality and technological prowess. China may use Soviet-era technologies considered outmoded by the U.S., but they also engage in electronic warfare with hackers and electromagnetic pulse weapons. The GLA's lack of material resources is balanced with a higher degree of

17 Another crucial part of the RTS formula is the feeling of constantly assaulting an opponent in a cruel and unforgiving manner. Between these two feelings, the latter is generally considered preferable.

versatility within unit types, and only the GLA possesses the practical ability to upgrade their own units with the remains of destroyed enemy vehicles. Similarly, while the Chinese and GLA can be seen as treating their own troops as disposable, this is only a matter of degree: the battles are large enough that no army can be allowed the luxury of worrying about the survival of any individual unit. It is simply unavoidable, assuming a reasonably competent opponent, that many of the players' troops are explicitly being sent to their deaths.

This is, of course, a necessary part of warfare as it is currently defined. "Collateral damage" in the form of civilian casualties is another. Kant does not dwell on the specifics of ethics in warfare, and utilitarianism offers only the vague guideline that actions on the battlefield are moral to the extent that they maximize happiness and minimize suffering. Delineating a theory of moral warfare according to either philosophy is a task I do not intend to undertake here. Rather, I intend to look to general aspects of the two philosophies, and the gameplay principles laid out in the previous chapter, to see how they could be applied to the treatment of civilians in military action against non-state actors, using *Generals* as a model.

Either philosophical perspective will assign some degree of negativity to the killing of civilian non-combatants. The central distinction is one of motivation: why do the characters in the game do what they do? If they are motivated by duty to the moral law for no reason save duty itself, some form of Kantianism will apply; if are they motivated by a desire to improve the sum utility of humankind, some form of utilitarianism will apply. The question then becomes, under what circumstances is it permissible to kill civilian non-combatants?

Assuming, for the sake of argument, that lethal violence against civilian non-combatants cannot be reconciled with Kant's moral law, intentional targeting of said non-combatants cannot be moral, according to Kantian principles. Even if the option is available to the player—and even if it is useful, from a military perspective, to do so—intentional violence against non-combatants must have a negative, non-diegetic effect on the gameplay that is in some way connected to the win condition. Conversely, a utilitarian approach will allow the player long-term benefit from such tactics, as long as these civilian deaths can be shown to save lives in the long run. In practice, a strict Kantian approach (as defined here) would be very simple, and a strict utilitarian approach would prove very complicated. Beyond that intentional, lethal violence must be directed only at those explicitly identified as combatants, there is little more to be said about the Kantian approach. The utilitarian approach would require a way to quantify the long-term consequences of a given battle: the overall happiness of a given territory, of course, should vary enormously based on which side wins. Assuming one side is morally superior to another—and the game must make this apparent through gameplay—any evil done on the battlefield should prove negligible in the wider scope of history. Even from a utilitarian perspective, however, there might be consequences to such a scorched-earth battle strategy. *September 12th* presents one such consequence, suggesting that killing non-combatants causes the number of active combatants to increase. This principle could easily be expanded, for the GLA has other unit types requiring little in the way of specialized combat training.

But what constitutes a civilian? While the GLA does have a military command structure that functions similarly to that of the U.S. and Chinese forces, certain units seem

to be outside this structure. Civilian non-combatants litter the streets in urban maps, and provide no threat in and of themselves. Yet, these civilians provide cover for certain units, and provide "raw materials" to grow angry mobs. These civilians are clearly not combatants before they join a mob, but they are equally clearly combatants afterward. If the player cannot discern which civilians will become violent, to what extent can players justify killing civilians that *might* become a military threat later in the game? To what extent can workers, who provide material support to the GLA but are not themselves a threat, be targeted? Furthermore, how quickly will the costs of such action (in the form of increased opposition) outweigh the benefits?

Neither of the philosophical approaches outlined in the previous chapter are especially popular in terms of current beliefs about the type of conflict represented in *Generals.* Most current popular opinions draw from both, seeking a balance between the long-term utilitarian benefits of military victory and the avoidance of war crimes in the short-term. Assuming the designers do not wish to explicitly align the game's morality with a specific philosopher, a number of arguments are available, based on whether non-empirical consequences (i.e. violations of the moral law) are counted, and how quickly the costs of targeting civilians outweigh the benefits. These are largely the same issues that must come into play in any discussion of another key aspect of the "War on Terror," the morality of torture.

Four recent, commercial games have directly dealt with the issue of torture: *The Punisher, State of Emergency 2, The Godfather* and *Reservoir Dogs.* This list is not exhaustive, but these titles demonstrate some of the ways torture has been approached in existing games. Of these four titles, *The Punisher* is the most explicit, and is the central

subject of my investigation. As such it receives the most attention, but all four offer useful insight on the subject.

The Punisher, it must be noted, is not merely a game, but part of a multimedia franchise. Originating as a villain in an issue of *Spider-Man,* the character known as Frank Castle—alias The Punisher—has been a persistent figure in the Marvel Comics universe for thirty years. Volition's videogame adaptation of *The Punisher* was released in 2004 to coincide with the theatrical release of the film of the same name. Both the film and game adaptations drew heavily on the work of Garth Ennis, who had recently revitalized interest in the character among comic readers. Ennis' particular take on The Punisher is substantially more complex than the simple-minded vigilante previous writers had crafted, and the *Punisher* videogame is so thoroughly steeped in the work of Ennis that it cannot be read in isolation from that work. Panels from Ennis' books provide a substantial part of the game's reward system, and serve as indexes, pointing to the larger narrative of which the game is a part. That narrative guides the game mechanics, and the game's ethical framework compels the player to kill in a variety of ways, none of which should be unfamiliar, symbolically or mechanically, to any action game enthusiast. What is comparatively new is *The Punisher*'s treatment of torture.

The Punisher's so-called "torture engine" is a mini-game of sorts. Frank puts his victims in a dangerous, frightening and/or painful situation that is not immediately lethal, and he must keep them sufficiently intimidated without killing them. The controls vary with every method of torture, but all rely on subtle manipulation of an analogue stick. At first glance, torture appears to function as an interrogation technique. Certain characters possess special information that can only be extracted through torture. However, this

information is never essential to Frank's mission, but only supplementary: a skilled player can easily get by without it. Moreover, very few characters have *any* useful information to be extracted, yet nearly all can be tortured. In spite of torture's lack of value for interrogatory purposes, it is nevertheless a crucial play mechanic, and players cannot easily avoid engaging in it.

The Punisher is not an open-ended play-space like *Second Life,* and players are not expected to do things merely because they can. Rather, the game's ethics encourage torture by connecting it to two incentives: the acquisition of points, and the unlocking of hidden content. Points feed directly back into the gameplay experience, as players exchange them for skill and weapon upgrades. Scripted, location-sensitive tortures provide the largest point bonuses, but any enemy character within grabbing distance can be exploited for this purpose, and an execution is never as profitable, in terms of points, as an execution preceded by torture. In addition to the points, torture will randomly cause Frank to have flashbacks. These flashbacks are presented to the player as a panel of comic art from Ennis' *Punisher* stories accompanied by an appropriate voice sample: for example, an image of Frank holding a dead family member juxtaposed with a terrified criminal screaming "I have a family!" These flashbacks, once unlocked in the main game, can be viewed from the title menu, and contribute to overall completion of the game, much like the side-quests in the recent *Grand Theft Auto* games. For the player, the reward for the (frequently challenging) act of torture is non-diegetic. Points have no meaning at the narrative level, and it's unclear why Frank would want to suffer flashbacks to painful moments in his life. Thus, in terms of the game's internal world, it would be tempting to refer back to George Orwell's *1984*: "The purpose of torture is torture."

More accurately, though, the purpose of torture, in *The Punisher,* is a "bonus round" of sorts, a chance to allow the player to demonstrate skill in exchange for points. If torture is a "mini-game," it is easy enough to "fail" by accidentally killing the victim. The player loses points for killing a victim in the course of torture, even though he or she would *gain* points for killing the same person in a more conventional fashion. The game takes no notice whether or not the victim has given Frank whatever information they have. The rules are simply that killing is rewarded, torture is rewarded, but accidental killing *during* torture is punished. These are the ethics of torture in *The Punisher,* and they make sense at a purely mechanical level. At a narrative level, they are internally inconsistent, and thus the narrative and ethics cannot be integrated into a moral argument.

State of Emergency 2 is the little-known sequel to the controversial *State of Emergency,* which places players in violent street combat against a fascistic corporate dictatorship. The original game incorporates contemporary political debates about globalization into its narrative, but squanders its potential for legitimate discourse through simple-minded play mechanics.

The sequel adopts a more linear, story-based approach to revolution that includes a mini-game in which players interrogate suspects. The interrogator is "Spanky," a former gang member and Hispanic stereotype, and the interrogation consists of repeatedly punching a captive. In terms of play mechanics, interrogation is a timing game, in which players must hold the proper button and release it at the proper time—release the button too early and Spanky will not punch hard enough to cause sufficient pain, release the button too late and Spanky will punch too hard and kill the captive. In contrast to the calculated brutality of the torture seen in *The Punisher,* the *State of*

Emergency torture scenes are somewhat cartoonish. The famously graphic violence of the original *State of Emergency,* which allows players to blast NPCs apart with explosives and then use the charred body parts as weapons, has been toned down considerably in the sequel, and one wonders why torture was included at all if gratuitous violence were a concern. As it stands, the torture scenes are among the *least* violent and disturbing action scenes in the game.

The Godfather is the high-profile videogame adaptation of the world described in the Mario Puzo novel and Francis Ford Coppola films. Though not explicitly mirroring the plot of the novel or films—the protagonist is a new character not found in either—the ubiquity of *The Godfather* in popular culture makes it unlikely that players will come to the game unfamiliar with the Corleone dynasty. As with *The Punisher,* the game narrative must be read in context of the larger text of which it is a part.

Intimidation is a major factor in the gameplay of *The Godfather.* The most common use of intimidation is against shopkeepers, to encourage them to hand over protection money. Unlike the previous examples, the player need not resort to physical pain for this purpose, although the game allows a great deal of realistic physical violence. If a shopkeeper is being particularly stubborn in his refusal to pay, smashing his cash register might be more effective than choking him or shooting him in the kneecap. Consistent with the gangster ethics detailed in the novel and films, the game engine generally rewards players for finding ways to intimidate without resorting to direct bodily harm.

Finally, *Reservoir Dogs* is the videogame adaptation of the 1994 Quentin Tarantino film of the same name. Similar to *The Godfather,* torture is used not for

interrogation, but rather for intimidation. Though the game gives players the option of blasting their way through all obstacles, earning a "Psychopath" rating in the process, the more cerebral "Professional" track requires a more measured use of violence, both threatened and enacted. Taking human shields, and therefore threatening hostages with lethal violence, is sufficient to disarm security guards, but will result in a standoff with actual police. Police will also drop their weapons, however, if the player pistol-whips the hostage in front of them—but even this is ineffective against large numbers of police. When surrounded, players who have charged up the avatar's "adrenaline" can perform a "signature" move, beating the hostage into unconsciousness and likely disfiguring him or her in the process.

These "signature" moves are unique to each character, from Mr. Blue's cigar to Mr. Blonde's trademark straight razor, though the most brutal violence happens off-screen. A "signature" move will make every cop in the vicinity lay down their weapons in surrender. The game's ethics, in this case, cannot possibly be developed into a moral argument, simply because the they *make no sense whatsoever at the narrative level.* Beating and disfiguring a civilian should, logically, make the character more likely to be shot by police, not less. In addition, unconscious hostages drop to the ground and cannot be picked up. Thus, by performing a "signature" move, the protagonist reveals to the police that he is violent, unpredictable and dangerous, while simultaneously releasing his human shield. The torture techniques described by Mr. White in the film, or enacted by Mr. Blonde, would have made some degree of sense, but the torture found in the game, while superficially similar, does not.

In all these games, some common elements exist. First, the games' ethics, which

compel the player to torture, are not explicitly out of sync with the protagonists' motivations. From the protagonists' perspective, torture is justified by the moral "gray area" of the situations in which they find themselves, be it organized crime, insurrection, or simple mass murder. We are given no reason to believe that the protagonists themselves believe torture to be immoral, at least under the given circumstances. It is worth noting that three of the games I've discussed, *The Punisher, The Godfather* and *Reservoir Dogs,* are adaptations of existing works, and each inherits a nuanced morality of violence from the worlds' origins in film, novels and comic books. The player is not called upon to accept or reject the protagonist's actions as moral, and the circumstances in which the protagonists find themselves are defined as extraordinary and largely unrelated to "real life."

Second, the morality of torturing an innocent is never addressed. The Punisher cannot torture an innocent person who was simply in the wrong place at the wrong time, because these people do not exist in the game.[18] In *The Godfather* and *Reservoir Dogs,* the player is an anti-hero at best, but there are no judgments on when it is *moral* to torture, just when it is ethical in terms of gameplay.

Third, when torture is applied for the purpose of interrogation, it is universally effective. The tortured party will invariably "crack," given the right circumstances. When they do, they will invariably give the protagonist correct information.

Fourth, the actions of the player have no long-term effect on the overall "war effort." It is hard to imagine how it could, given the genres in which it takes place. The

18 Innocents do exist in *The Punisher,* but they are clearly identified as such and Frank cannot be mistaken about this fact.

mafia and the fascist thugs of the games in question are not in a position to become *more* brutal due to the avatar's actions.

Fifth, the experience of having intentionally inflicted pain on a defenseless human being has no long-term effect on the mental health of the protagonist. Again, this is to be expected, since the modeling of avatar's mental states is still very rare in videogames.[19]

These games clearly demonstrate that videogame designers have developed the conceptual tools necessary to model the act of torture, but not its consequences. By carefully integrating the rule system and narrative, and by explicitly addressing those elements found lacking in the games I've described, it is possible to design videogames that make coherent moral arguments about, and more specifically against, torture in a way that would not be possible in any other medium. I here propose a model for such a game.

The best genre for such a game would be a single-player strategy game that alternates between macro-management and micro-management, similar to Microprose's *X-COM: UFO Defense.* Time will need to be somewhat fluid in the game, which would suggest a turn-based approach, but there's no reason parts of the game couldn't be designed for real-time strategy, using the modified *Command & Conquer: Generals* engine described earlier. The player commands a military unit in occupied territory, under constant threat of attack from local guerrilla forces. To prepare for or prevent these attacks, the player must gather information, make arrests, interrogate suspects, and use the new information to coordinate attacks or make more arrests. Like *X-COM,* gameplay

19 *The Sims* makes some progress towards this goal, as do certain titles in the "survival horror" genre such as *Eternal Darkness, The Suffering* and the *Silent Hill.*

will be cyclical in nature, and will end when either the guerrillas successfully wipe out the player's unit, or when public support for the guerrillas wanes and order is restored. These are only end conditions, however—it might be necessary, depending on the argument the designers seek to make, for true, non-diegetic victory to be independent of military success. Most importantly, the morality espoused in the narrative must be consistent with the ethics of gameplay.

As the game begins, players are given some initial intelligence from a variety of sources concerning planned attacks, and suggesting suspects. Players must then travel to a given location and attempt to arrest a suspect, using a minimum of force. After all, killing a suspect before he can make himself useful is a failure at both military and moral levels. Assuming the suspect can be arrested and returned to base successfully, the interrogation phase begins.

The interrogation process is the most significant portion of the game. Consequently, the game rules must acknowledge the issues ignored by the games I've discussed. The rule system, after all, will determine the ethics of gameplay, compelling gamers to play in a certain way, and the narrative cannot be allowed to disconnect from these ethics. Thus, characters must express differing opinions on the morality of torture in general. Establishing the opinions of NPCs can be handled in a number of ways, and designers need not resort to overlong cutscenes, but they will need, at the very least, well-written dialogue that is both semi-random and likely to be encountered by players. In addition, the game must include the possibility of bad intelligence, and it must be possible, even likely, for players to make false arrests. Whether or not the suspects actually know anything, many will lie and give false information as the torture becomes

increasingly brutal; conversely, some will protest their innocence through any level of torture, and some will simply say nothing.

Players will be allowed to detain suspects for as long as they choose, torture them in any way provided by the game designers, and execute them at will. All of these actions must directly affect the rest of the game. The guerrillas might gain popular support, and become more numerous and better armed, depending on who the player arrests, how the suspects are treated, and whether they are released, detained indefinitely, or executed. In addition, as a result of the player's actions, suspects could become increasingly less likely to allow themselves to be arrested, opting instead to shoot it out with the player's troops or blow themselves up to evade capture.

In addition to the effects of the player's torture on the effectiveness of the mission, there must also be consequences to the torturer. This can best be accomplished by having a single interrogation specialist character with greater narrative depth than most other characters: in the context of the interrogation sequences, the specialist is the protagonist. While much of the game's dialogue can be semi-random, the interrogation specialist must have more tightly scripted dialogue, and more of it. If the game is to have a narrator of any kind, the interrogation specialist is the logical choice. As torture becomes more frequent and more brutal, the specialist will become increasingly unhinged. Torture will become more difficult to accomplish, as the protagonist increasingly "ignores" the player's controller input, increasing the number of so-called "accidents." As the protagonist moves from torture as a means to an end to torture as an end unto itself, he will become less effective at extracting information. The less brutal methods of interrogation will cease to be available to players. Eventually, it will become impossible

for players to do anything with suspects *except* brutally torture and kill them, and doing so will only hasten the victory of the guerrillas.

These are the basics of the game, the elements common to any meaningful argument against torture. From there, three specific arguments can be made. The specific mechanics of the game, such as the probabilities of arresting an innocent person or extracting false confessions, will be dependent on the designers' intended argument. The first is a rather Machiavellian claim that torture is an effective tool for a counter-insurgency, but must be used sparingly, so the benefits of useful information outweigh the costs of increased enemy resistance and deaths of innocent victims. This argument defines what is good as what wins the war, and treats torture as an evil to be engaged in only for a greater good. For this argument, torture must make the game easier to complete; refraining from torture as much as possible must bring a greater difficulty and a greater reward. Nonetheless, the only win condition is military victory, and no moral rule is more important than that.

The second argument is that torture is simply counter-productive. For this argument, the variables must be set so the costs of torture are overwhelmingly larger than any possible benefits. Consequently, it must be impossible to complete the mission using torture as a strategy, and victory must be easiest when the player repudiates torture entirely. Again, this argument ties morality with military victory, and the most moral solution is also the most practical. This argument could also be made satirically by separating the win condition from military victory, and rewarding the player in non-diegetic ways for continuing to torture even as it dehumanizes the protagonist, kills innocent people, and allows the guerrillas to take over the country. The world will be

decisively worse than when the player began the game, the mission will have failed miserably, but the player will be assured, through a high score, that they've done the right thing. The sheer absurdity of such a game would be a powerful argument against torture.

The third argument differs from the first two by designing the game's ethics to serve an anti-torture morality completely divorced from military victory. The mission may succeed or fail, but such success is not taken into consideration in terms of the player's reward. Rather, the game must encourage players to torture by offering powerful short-term benefits, and reward them for resisting the temptation, both with non-diegetic rewards such as points and unlocked content, and a well-constructed narrative that makes it clear that, win or lose, soldiers who refrain from crimes against humanity can at least look themselves in the mirror with their sanity intact.

Conclusion

I have here laid out one possible vision of what morally meaningful videogames might look like. My approach is drawn from a wide variety of videogame theorists, as well as my own experiences as a player, as a consumer of many forms of media, and as a writer. I have attempted to define a somewhat holistic view of videogames in which recreation of other media forms is an inherent feature of a medium that need bias neither narrative nor simulation, but can present a unique synthesis of the two. The videogame medium comprises an enormous variety of texts, and these texts do not all exploit the same elements of the medium. Videogames, as a whole, are simulations *and* stories, as well as sports, spaces, and games. Any given text will use these elements differently.

Given this array of possible approaches, my argument is applicable only to a subset of videogame texts: single-player games with explicit narratives, explicit win conditions and moderate degrees of linearity. The articulation of a moral argument through the synthesis of gameplay ethics and narrative requires that such a narrative exist, and that the designers must intend the player to complete the game in a certain way—a way whose meaning is defined as moral by the narrative. I include the "single-player" qualification primarily as a measure of control, as any multi-player game invariably results in the creation of "local rules" that cannot be reasonably enforced, and might explicitly contradict the designers' intent. The designers' intent is not sacrosanct, of course, but my concern in this paper has been primarily with authors, not readers.

The readers, of course, present some complicating factors. Designers' power to shape their games only extends so far. Players can assign complex narratives to games in which they were not written, and can just as easily ignore the narratives that exist.

Moreover, my argument does not apply to metaplay, in which players informally alter a game's rules, overriding the canonical design. For players to experience and interpret the designers' argument in such a game, they must play it the way it was intended to be played at least *once,* which not all will do. Even in stealth-themed games like *No One Lives Forever,* many players choose to play the game as a conventional first-person shooter and rack up the highest possible body count. Designers cannot, ultimately, *force* players to obey the gameplay's ethical imperatives, any more than they can force players to pay attention to the story. In the Kantian *Fable,* for example, it is necessary for a player to act on Kantian principles to achieve the best ending, and it would certainly be useful to understand those principles on a theoretical level. Nonetheless, it would be easy enough for a player to find a walkthrough and guide the protagonist toward moral perfection without understanding why his actions were connected to morality.

In short, although videogames can make arguments about morality, they are not a magic bullet that circumvents the usual intellectual process. Videogames are texts, and their communicative properties are limited to those of texts. However, texts can be powerful things, and any medium that possesses the ability to argue or persuade can theoretically become an instrument of propaganda. In "The War between Effects and Meanings," Henry Jenkins describes two opposed models of how videogames might function as teaching machines: "effects" and "meanings." The effects model, espoused by figures such as David Grossman, argues that games teach by a simple stimulus/response mechanism:

> Grossman assumes almost no conscious cognitive activity on the part of the gamers, who have all of the self-consciousness of Pavlov's dogs. [...] Grossman sees games as shaping our reflexes, our impulses, our emotions,

> almost without regard to our previous knowledge and experience. ("Effects and Meanings" 211)

The meanings model, conversely, acknowledges that players construct their experience of a game from a variety of sources, many of which hinge on the players' own sense of identity, self-defined goals, and the "previous knowledge and experience" Grossman wishes to circumvent. Jenkins looks to his experiences in the classroom for insight into problems with the idea of purely unconscious transmission of texts:

> As a teacher, I may fantasize about being able to decide exactly what I want my students to know and to transmit that information to them with sufficient skill and precision so that every student in the room learns exactly what I want. But real-world education doesn't work that way. Each student pays attention to some parts of the lesson and ignores or forgets others. Each has their own motivations for learning. Previous understandings and experiences color how they interpret my words. Some students may disregard my words altogether. There is a huge difference between education and indoctrination. ("Effects and Meanings" 212)

It is ironic that, for all the concerns about videogames interfering with education, the effects model seems to assume that videogames are better teachers than human beings. The types of games I have described here are not explicitly designed for classroom use, but I think of their function as being inherently educational, in the sense that any text sufficiently complex will have educational value by encouraging critical thinking. Their ability to function as propaganda is limited by both the tendency of players to "tune out" elements of the game that they find uninteresting, as well as experienced gamers' hard-won ability to critically read videogame texts.

The potential of videogames to make moral arguments is both exciting and disconcerting, as there is no shortage of governments or corporations that would enthusiastically embrace a new channel for propaganda. I have described games that

argue against the moral legitimacy of torture, but there is no reason why these principles could not be used to design a game that argued the moral *necessity* of torture. The time and resources involved in videogame production ensure that it will be easier for some groups than others to make or finance this kind of game. If morally meaningful videogames are to be a reality, lack of equal access to the technology presents an alarming problem. However, the solution to this problem already exists. Thanks to open-source coding and off-the-shelf design tools, amateur-designed games are gaining prominence. Any group that seeks to use videogames as propaganda must accept the possibility that amateurs could mod the game and dramatically change its message. As always, the solution to speech is speech.

I am not claiming that videogames will replace the existing methods by which we articulate, debate, and teach moral philosophy. Although the games I have described here might be useful supplements to students of philosophy, I would not expect the playing of the Kantian *Fable* to replace the reading of Kant anymore than I would have expected *The Screwtape Letters* to replace the Bible. What I am proposing is a new way to look at videogames, and a new way to think about philosophy.

As a player and a theorist, thinking about videogames in terms of moral arguments led me to rethink a number of my preconceptions about how ideas are communicated and how the avatar functions in modern videogames. The distinction between morals and ethics arose from examining how games tend to conflate the goals of the protagonist (a character in a story) with the goals of the player (someone playing a game), and how the results were often morally questionable. This distinction between what the player needs to do to win the game and what the protagonist ought to do as a

sentient being in a certain situation—the ethical imperative and moral imperative, respectively—necessitated a more nuanced view of the avatar. My earlier separation of the avatar into "player" and "protagonist," and my distinction between morals and ethics in videogames, are two facets of the same idea: that the worlds described in videogames can be imagined to exist even where there is not a player, and that an avatar's actions can (but do not always) mean different things to the player and the protagonist. Videogames are not films, nor are they books, but the characters *in* the games—on their own terms, without taking the player into account—can be interpreted as such, to help delineate differences between the perspectives of the player and the protagonist. This perspective has greatly enhanced my experiences with videogames, allowing me deeper and more meaningful play experiences and allowing me to think critically about them in new ways.

As for philosophy itself, videogames offer an intriguing opportunity to engage with it in a way that is both intuitive and tactile. Were any of the games I've described to be built, play would certainly reveal areas where the designers' application of the philosophical principles failed to adequately reflect the philosophers' views, or where the principles appeared to contradict one another. In either case, play would encourage a deeper understanding of the texts in question; in the latter case, players would need to determine whether the incongruity was the fault of the designers, or the philosophy itself.

I believe the potential for videogames to model and critique elements of moral philosophy is vast. I have delineated only a few examples here, but there is much more work to be done. My work here has dealt only with a small subset of videogame genres, and other genres might lend themselves to entirely different types of arguments, or entirely different approaches to the kinds of arguments I've discussed. The videogame

medium, reduced to component parts, offers little that is new, but in synthesizing these disparate elements, it offers a new way to communicate. A moral perspective posits a certain kind of world, and videogame designers have been building worlds from the beginning.

Works Cited

Bentham, Jeremy. *An Introduction to the Principles of Morals and Legislation.* New York: Hafner Press, 1948.

DC Studios. *State of Emergency 2.* PlayStation 2. Southpeak Interactive, 2006.

Electronic Arts. *Command & Conquer: Generals.* PC. Electronic Arts, 2003.

___. *The Godfather.* PlayStation 2. Electronic Arts, 2006.

Favaro, Peter J. *Alter Ego.* PC. Activision, 1986. Emulated online at: <http://www.theblackforge.net/>

Frasca, Gonzalo. "Simulation versus Narrative: Introduction to Ludology." *The Video Game Theory Reader.* Eds. Mark J.P. Wolf and Bernard Perron. New York: Routledge, 2003. 221-236.

___. *Videogames of the Oppressed.*

Gee, James Paul. *What Video Games Have to Teach Us About Learning and Literacy.* New York: Palgrave Macmillan, 2003.

Gregor, Mary J. *The Cambridge Edition of the Works of Immanuel Kant: Practical Philosophy.* New York: Cambridge UP, 2006.

Grossman, David. "On Killing II: The Psychological Cost of Learning to Kill." *Killology Research Group.* 8 May 2007. <http://www.killology.com/article_onkilling.htm>

___. "Trained to Kill: Are We Conditioning Our Children to Commit Murder?" *Christianity Today.* 10 Aug. 1998. 8 May 2007. <http://www.killology.com/article_trainedtokill.htm>

Huizinga, Johan. *Homo Ludens.* Boston: Beacon Press, 1971.

Jenkins, Henry. “Game Design as Narrative Architecture.” *The Game Design Reader.* Eds. Katie Salen and Eric Zimmerman. Cambridge: MIT Press, 2006. 670-689.

___. “The War between Effects and Meanings.” *Fans, Bloggers and Gamers.* New York: New York UP. 208-221.

Kant, Immanuel. *Critique of Practical Reason.* Gregor 133-272.

___. *Groundwork of the Metaphysics of Morals.* Gregor 37-108.

___. *The Metaphysics of Morals.* Gregor 353-604.

___. *On a Supposed Right to Lie from Philanthropy.* Gregor 605-616.

Krzywinska, Tanya. “Hands-On Horror.” *ScreenPlay: cinema/videogames/interfaces.* Eds. Geoff King and Tanya Krzywinska. London: Wallflower Press, 2002. 206-223.

Lionhead Studios. *Black & White.* PC. Electronic Arts, 2001.

Lionsgate. *Reservoir Dogs.* PC. Eidos Interactive, 2006.

LucasArts. *Star Wars: Dark Forces II: Jedi Knight.* PC. LucasArts, 1997.

Mayer, Jane. “Whatever It Takes.” *The New Yorker.* 19 Feb., 2007.

Microsoft. *Fable: The Lost Chapters.* PC. Microsoft, 2005.

Mill, John Stuart. “Utilitarianism.” *Utilitarianism, On Liberty, and Considerations on Representative Government.* Ed. H.B. Acton. New York: E.P. Dutton, 1977. 1-64.

Poole, Stephen. *Trigger Happy: The Inner Life of Videogames.* London: Fourth Estate, 2000.

Rachels, James. *The Elements of Moral Philosophy.* 3rd ed. McGraw-Hill, 1999.

Rauch, Peter. “Free Will in a Predestined World.” Unpublished Massachusetts Institute

of Technology term paper. 12 Oct., 2006.

___. *Videogames as Protected Speech.* B.A. Thesis, Florida Atlantic University, 2004.

Troika Games. *Arcanum: Of Steamworks and Magick Obscura.* PC. Sierra Entertainment, 2001.

Volition. *The Punisher.* PlayStation 2. Capcom, 2004.