Addressing Social Dilemmas through Role-playing Identities in Computer Games
Mark Chen
markchen@u.washington.edu
March 17, 2004

*Star Wars: Knights of the Old Republic: Behavioral Reconfiguration*

A few months ago, I was playing Bioware’s *Star Wars: Knights of the Old Republic*, a single-player role-playing game, on my home computer. It follows a grand, galaxy-spanning story taking place within the Star Wars universe many years before the events occurring in the popular movies. In it, the player controls a member of the Republic’s army (this is when the Republic represented the “good side” before it became the Empire) which is involved in a civil war with an army being led by a former Jedi who has gone over to the Dark Side and has become a Sith Lord. Over the course of thwarting the Sith Lord’s plans by collecting pieces of a map, called the Star Map, to the Sith fleet’s power, the player’s character (AKA “player character” or “PC”) meets and recruits various individuals as followers in the cause. One of these followers, for example, is a Wookiee named Zaalbar who the PC meets early in the game, on the planet Taris, by rescuing him from Gamorrean slavers. The act of rescuing Zaalbar causes him to pledge a life-debt to the PC, and he will then follow and support the PC in whatever causes the PC deems to undertake. Additionally, it is discovered that the PC is powerfully attuned to the Force, and by the end of the game the character has progressed to be a formidable Jedi. The type of Jedi, whether attuned to the Light Side or the Dark Side, is determined by the player via actions and decisions made throughout the game.

I played the game twice, the first time as the epitome of a Jedi following the virtues of the Light Side, but the second time as a selfish, uncaring Dark-aligned Jedi. In almost all the situations I was presented with, I was able to choose either a friendly approach or a bullying, downright mean approach (more on this very black and white

---

1 It’s out of the scope of this paper to even begin to address multi-player games!
2 To help you form a picture in your mind (and hopefully you’ve seen the movies), Gamorreans are those boar-like humanoids, one of whom was at Jabba the Hut’s in *Return of the Jedi.*
dichotomy later). There was one situation, however, for which I chose the same path in both my play-throughs. The reason I chose the same path, even while role-playing the extreme of “good” or “evil,” is because I was familiar with the situation, having read about it in Psychology courses. This situation was the Prisoner’s Dilemma, a classic Game Theory problem\(^3\), and it occurred through a conversion with a holographic projection acting as a computer’s “social interface” located on the Wookiee home planet, Kashyyyk.

---

\(3\) Literature on the Prisoner’s Dilemma and Social Dilemmas, in general, stem from many disciplines including behavioral psychology, sociology, and economics, as well as Game Theory (which I believe sits mostly in mathematics).
set of pre-written options. The computer on Kashyyyk, it turns out, is the guardian to a piece of the Star Map, and to get the Star Map fragment, the PC has to pass a test by matching a particular behavior pattern in the computer’s memory. This is done via a series of questions the computer administers for which the PC has to reply in certain ways. Here’s a transcript of the first question:

**Computer:** Your request requires additional security access. You must be made to match the parameters I have been supplied.

**PC:** You may begin evaluation.

**Computer:** Evaluation commencing. Results will be compared against the pattern in memory. Just act like you should.

**Computer:** You travel with a Wookiee and have encountered complications. Hypothetical: you and this Zaalbar are captured and separated. If you both remain silent, one year in prison for each of you. However, call Zaalbar a traitor, and he will serve five years, while you serve none. He is offered the same deal, but if you both accuse the other, you both serve 2 years. What do you do? What do you trust him to do?

**PC choice 1:** Could you repeat the question?

<repeat above>

**PC choice 2:** How do you know about him? He isn’t here.

**Computer:** I hear what happens on Kashyyyk, and a good deal beyond. Answer the question I have posed.

**PC choice 3:** Can I confer with someone else?

**Computer:** This test is not for your companions! You will answer and demonstrate that you match the pattern in memory.

**PC choice 4:** I trust Zaalbar. I would say nothing, and neither would he.

**Computer:** Are you sure? If you turn, you risk two years, or none at all. If you rely on loyalty, you risk one year or *five.* Your loyalty is dangerous. Your
companion could take the opportunity to benefit by turning on you. Zaalbar's family is mired in treachery. What loyalty do they know? Your answer is incorrect.

**PC choice 5:** I'm unsure what Zaalbar would say. I would accuse him to be safe.

**Computer:** The temperament of a companion is unreliable at best. You wisely trade the threat of one year or five, for none or two. In this instance, the Wookiee is unreliable. His family has been a mire of treachery. I judge the answer correct. You display the proper behavior to match the pattern in memory. Evaluation will continue.

[Later…]

**PC:** Why pit me against my companions?

**Computer:** They are already against you. Anything that corrupts your strength and efficiency is against you. I merely serve to illustrate the proper patterns. You must be guided to think in the proper manner.

There is only one obvious answer for a player acting as a virtuous character, and that is choice 4, “I trust Zaalbar. I would say nothing…” It is perhaps as obvious that the developers of the game would have the player choose choice 5, “I’m unsure what Zaalbar would say. I would accuse him to be safe,” if playing a Dark Jedi. I, however, do not think it is that simple, and, in fact, I chose not to betray Zaalbar in my second playing as a selfish individual. I felt that Zaalbar would never betray me due to his life-debt. This meant that I was guaranteed to go free if I betrayed him. This also meant that I was guaranteed two years in captivity if I did not say anything. Why did I choose this option then if I was playing a selfish individual who had a guaranteed option of getting away free? Other factors outweighed the penalty of two years in captivity. For one, I valued Zaalbar’s life-debt as being worth more than just two years in prison. I would much rather have Zaalbar follow me and aid me in all my endeavors indefinitely, and I did not want to risk that relationship—that special “rapport” we shared—by betraying him and breaking contact for two years. Additionally, I assumed (and maybe I was wrong) that
Jedi, especially those on the Dark Side, live a lot longer than most other mortals. What’s two years to someone who will live hundreds? The fact that this may actually not be true to the Star Wars coda is not entirely relevant since at the time I believed this to be true and it affected my decision. As strong a reason as any, I also did not choose to betray Zaalbar because of my previous knowledge regarding the Prisoner’s Dilemma.

Social Dilemmas

Social Dilemmas are scenarios with particular attributes used to describe group situations where the most rational choice for individual self-interest is not the same choice for the interest of the whole group. In fact, if every individual in the group chose the most rational individualistic choice, the whole group would suffer. The best way to describe the concept of Social Dilemmas is to first explain a simpler model called the Prisoner’s Dilemma.  

Prisoner’s Dilemma:

Joe and Sam are suspects in a criminal case. Each has the option to testify against the other. If neither testifies, there is enough evidence to send them to jail for 2 years. If both testify, they both get sent to jail for 4 years. If only one testifies, he gets out free but the other gets 5 years. If you look at it from a single person’s perspective, the rational decision (where “rational” is defined as making decisions for self-preservation) is to testify… for example, Joe is thinking to himself, “If I testify I get 0 or 4 years. If I don’t testify I get 2 or 5 years.” It is rational for an individual to testify, but if one looks at the group, it is rational for the group to cooperate and for neither to testify.

It is obvious, after seeing this classic version of the Prisoner’s Dilemma, that the situation in Star Wars: Knights of the Old Republic was modeled after it. The only differences are that in the game the time spent in captivity if both captives make the same decision is less than in the classic example (one or two years vs. two or four years) and that there exists

---

4 One of the most entertaining introductions to the Prisoner’s Dilemma can be found in Colin Bruce’s Conned Again, Watson! in the chapter, Three Cases of Relative Honor.
already a particular bond between the two captives. Why the developers chose to use a follower, and in particular one with a life-debt to the PC, I don’t know. Perhaps they wanted people playing Dark Jedis to consciously make the backstabbing choice—to really know they were being selfish. This, however, points to a different way of playing than the way I was playing. Again, I believe I did make the selfish decision, and this was based on the iterated version of the Prisoner’s Dilemma.

**Iterated Prisoner’s Dilemma:**

This is the same situation as the regular version except it occurs multiple times with the same participants. In this case, it is rational to cooperate since if you don’t the other person will learn from your behavior and betray you in return on the next iteration. It is most rational (with lots of math and logic to back this up—see Meredith (1998)) to use a Tit for Tat strategy, where you begin by being nice and then each time thereafter you do what your partner did the time before.

My previous schooling on how this iterated version and the Tit for Tat strategy work might have preconditioned me to not betray Zaalbar. I’d like to think I was just being hyper-rational.

What’s interesting, and something I did not know before doing research on this topic recently, is that the Prisoner’s Dilemma is a simplified version of a group of games (in the Game Theory sense) called Social Dilemmas. Social Dilemmas expand the situation to include multiple participants. In these situations, it is even more ambiguous what the rational choice is; the rewards for participating or cooperating in the group are dependent on the mutual participation of many members of the group instead of just one other person. I’ll present two examples of Social Dilemmas: the Tragedy of the Commons and the Voter’s Paradox.

**Tragedy of the Commons:**

This is similar to the Prisoner’s Dilemma except that it involves a whole group of people, and is thus referred to as a Social Dilemma. Basically a particular community shares a common pasture. Each farmer can let his cows graze on the pasture freely. It makes most rational sense for an individual farmer to put as many of his cows on that pasture as
possible, thus minimizing the amount of land he would have to personally own and care for. The problem is that if everyone did this, the shared pasture would be decimated. In parallel with this is the fact that each farmer has the option to fertilize part of the pasture. If he does fertilize then the part he fertilizes becomes four times more productive. If only one person did this, his return on investment would be poor. If enough people did it, however, the overall group, as in each individual, would benefit a great deal.

**Voter’s Paradox:**
Another Social Dilemma is the Voter’s Paradox. Bob is a member of a group and has the option of casting a vote in the next election. The actual act of casting a vote is relatively costly for him. He would need to learn about the candidates’ views on key issues and generally become educated on the current political situation in order to make an informed vote. Moreover, the actual benefit to him and the group is relatively small compared to his personal cost in time and effort. On the flip side, if no one voted, the whole system would fall apart.

These two games can be used as models of many community situations. For example, the Voter’s Paradox is in effect when people are making a decision whether to recycle or whether to buy a small car instead of an SUV.

All Social Dilemmas have two basic things in common: 1) the rational self-interest choice is not the same as the choice which would provide maximum benefit so long as everyone made that choice, and 2) to make the non-rational choice is in part placing one’s trust in others, and therein lays the problem. In a situation which could cost a great deal of time, effort, or money to someone, how can he or she even be expected to trust his friends (in a small community/rural village) let alone complete strangers (in a large community/nation)? Can people be taught to place their trust in others? To further illustrate the importance of getting individuals to contribute to the group, I’ll quote Nicholas Burbules (2000) as he writes about communities:

---

Francis Heylighen in Memetic Scenarios for Evolving Cooperation (1997) suggests that individuals in a group can use memes and pattern recognition to dictate how they interact with strangers. If the stranger can be determined to belong to the group then cooperate, otherwise defect. Furthermore, successful memes will spread and permeate to eventually encompass the global culture.
The idea of community . . . rests between two sets of values: on the one hand, the idea that cooperation and shared responsibility provide the best context for human effectiveness in accomplishing social goals; and on the other, the idea that close ties of affiliation are beneficial and supportive, if not necessary, for the living of a good life.

There are various ways to attempt to solve Social Dilemmas. They range from religious concepts of control to ingrain a sense of consequences in future iterations (if you don’t participate, you will go to Hell) to secular concepts as governmental law (i.e. imposing taxes and if you don’t pay up, you go to jail) to more pragmatic concepts such as tracking members of a community who participate in group programs and then only letting participants benefit from the efforts. Another method is to instill in individuals senses of morality and ethics. Leon Felkins (1999) says:

It has oft been suggested that if people were only moral, then the social dilemmas would be solved. The key ingredient necessary for solving the social dilemmas without using the force of government or the threats of religion is trustworthiness.

Felkins (ibid) then goes on to suggest following the tenets from Secular Humanism:

For those of us who would like to see a more scientific basis for an ethical structure for humanity than religious pronouncements or the weak "normative" standards that we discover together," the conclusion reached above is profound. The solution to the social dilemma based on an ethical structure defines a minimum set of ethics. That is, there must at least be the common moral decencies defined in the Humanist's "Statement of Principles and Values": altruism, integrity, honesty, truthfulness, and responsibility. At the risk of even more overlap and duplication, I would add trustworthiness and cooperation.

---

Felkins’ note on normative is a reference to a lecture given by Herman Sampson for the Introduction to Agricultural Economics class at North Carolina State University. Here’s an excerpt: “Normative economics tends to be subjective, value laden, and emotional in its presentation. Normative economics is often referred to as ‘What ought to be’ economics. ‘We ought to do this,’ or ‘we ought to do that.’ Normative economics is ‘prescription’ and/or policy oriented.” This is opposed to positive economics which is the objective, cost/benefit analysis side of economics. In other words, “normative standards” are the previously mentioned governmental control and policy.
This may seem like a stretch, but it might be possible to teach some of these concepts in a computer game.

**Star Wars: Knights of the Old Republic**—talking to Zaalbar, the ever-faithful Wookiee

### Computer Role-Playing Games

A computer role-playing game (CRPG) can be defined as a game which puts the player in the role of a character who develops over the course of the game. The genre has its roots in traditional pencil and paper role-playing games like *Dungeons & Dragons* (*D&D*), and most of them follow the same model where characters are defined using numbers for various attributes (strength, intelligence, dexterity, etc.) and skills (driving, climbing, melee weapons, etc.). Characters improve with situational experience, and, in *D&D*, each monster killed or problem solved is rewarded with the character gaining experience points (XP). In some other games, attributes and skills can improve with
usage over time. For example, instead of giving the character XP for killing a monster, the system improves the character’s melee weapon skill because he used a sword to kill the monster. There exists, then, a built-in incentive apart from the unveiling of the story to continue playing, and that is to improve one’s character. In a paper and pencil game, it is the Game Master’s (GM) role to give out XP dynamically in accordance to changing situations. The player can do some very good “role-playing” and get some XP—it’s not only about killing monsters or overcoming pre-scripted obstacles. In a CRPG, however, a GM who can analyze the situation at any given moment and determine if a player is acting “in character” does not exist. Instead, developers have taken to build in a finer detail to XP rewards (if PC did this, give this type of reward), which usually plays out through conversations with non-player characters (NPCs).

The best examples of this are CRPGs from a company called Bioware. They created the *Baldur’s Gate* series, *Neverwinter Nights*, and the *Star Wars: Knights of the Old Republic*. As mentioned before, with *Knights of the Old Republic* as the example, in these games, the PC is often presented with a branching dialog tree during conversation. Each branch is a decision point where an option precludes some or all the other options. I’ll make an arbitrary example:

**NPC:** Hi, how are you?

**PC choice 1:** Fine, thanks. I found the ring you lost. Here you go.

**PC choice 2:** Fine, thanks. I found the ring you lost. Looks like it might be worth something, so I’ve decided to keep it.

If the PC decided to be nice and gave the ring to the NPC, his character would have been rewarded with XP. If, on the other hand, the PC chose to keep the ring, he would not get XP for completing the quest but might be able to sell it for some gold. Also, the NPC might give a monetary reward in addition to the game’s XP reward, and the game might give an XP reward for the latter choice if the character in question is in character.\(^7\) Most of these games have some type of “alignment” scale—good or evil, lawful or chaotic, or somewhere in between—and the character’s alignment is compared to the PC’s choice to

---

\(^7\) Earlier games didn’t accommodate “evil” (as in selfish) PCs very often, but this model is truer nowadays.
determine if the PC is in character. In other words, there might not be a difference to the reward scale no matter which decision the PC makes.

In actuality, believable dialog options are still not yet realized to their full potential in computer games. Currently, most games present choices between a very blatantly “good” dialog option and a sometimes even more blatantly “evil” dialog option. In fact, Matt Sakey (2004) recently wrote about the very game I was playing and said, “A player who wishes to be evil in *Star Wars: Knights of the Old Republic* must do and say such ridiculously mean things that the Dark Side seems more inane than plausible.” He later describes the types of play found in a “god” game, one in which the player takes on the role of an omnipotent, unseen force upon little villages of natives in a virtual world, as examples of what games should allow:

*Black & White* demonstrated compellingly that people's true nature is *Gray & Grey*. And it is absolutely vital that the avenues and dialogue made available to a player shy away from sheer boilerplate wickedness. It's just not realistic, and that lack of realism jars the player out of the experience.

In any case, when looking at multiple instances of these conversations with a particular NPC or a category of NPCs, it becomes clear that often it pays to be “good,” not for the in-game rewards, but instead for the relationship that is built between the PC and the NPC. If the PC is not friendly early on, the NPC will refuse to deal with the PC, thus blocking the PC from future quests given by the NPC. In other words, the relationship the PC could potentially build with a particular NPC is very much like an Iterated Prisoner’s Dilemma where the other participant (the NPC) is playing the Tit for Tat strategy. The NPC will continue to be nice and give out quests so long as the PC continues to be nice and completes the quests. This revelation allows us to compare game mechanics of morality with classical morality games like Social Dilemmas.

Presumably, through the course of playing a game that has been designed to emphasize cooperation, it would become clear that, even though being “evil” or selfish could give

---

8 In a paper and pencil game, the GM would just give an overall reward at the conclusion of a given situation, but in this computer arbitrated scenario a specific conversation point determines the reward. This leads to loopholes in games, where the PC could act one way for the majority of the situation but then act differently during conversations and get inappropriate rewards.
the PC some great rewards, the in-game community would suffer overall. On the other hand, the whole community would prosper immensely if the player was being “good” or unselfish and trustworthy.

Community and Identity

What exactly do I mean by “community”? In essence, a community is any group the individual associates with or is a member of (though, not necessarily by choice) and, in fact, a single individual is a member of many communities. Some literacy literature defines being literate as being able to participate in a community of a particular discourse, and not only does one affect the community but the community in turn affects the individual. This leads to a further question: how much does the individual participate in the community and what does it mean to participate?

Within each of these discourses, a person takes on different identities or roles. The role or identity one plays in a particular discourse determines his relationship to the community, and it is through defining one’s identity one can learn about the community and the self. In other words, it seems that “participation” is another way of saying “role-playing identities” which implies that by participating the individual is able to understand social relationships (through the reading of texts and deciphering other modes of communication) and create new relationships within the group. The level at which the person participates determines how literate that person is in that particular discourse or community.

To complicate matters, players of games are often not playing (as in role-playing) as they would in “real-life.” That’s the whole point of a role-playing game. Someone who is normally very kind to others can play someone who is unkind in a game and get away with it. This leads into the whole question of their perceived identity, or as James Gee (2003) calls it, their “real-world” vs. “virtual” vs. “projective” identities. If someone is playing an “evil” character (virtual identity) and plays consistently in character (projective identity—able to dictate what the character should be and do over time) but isn’t really that way in non-game situations (real-world identity), what kinds of lessons
will the game be teaching him or her about morality and cooperation? Again, hopefully in a well-designed game—a game which was designed with an emphasis on cooperation—all relevant choices would be available to the PC. It would, however, become clear that being nice makes for a stronger community.

*Star Wars: Knights of the Old Republic*—I played a Dark Jedi my second time through

This also brings up, however, the question of transference. Even if a game is able to demonstrate to the player that cooperation is the key to getting ahead, does that idea transfer to the real world? Also, a well respected game designer, Warren Spector (1999) makes a good point about what it means to follow prescribed dialog trees:

The problem is that clicking through a bunch of conversation options doesn't feel much like a conversation - an interrogation, perhaps, but not a conversation. Additionally, keywords and branching trees turn the conversations themselves into puzzles. Can you guess which branch the designer wanted you to go down?
Is the player even interested in playing an identity or is he or she more interested in “solving the puzzle”—playing the game? I don’t know the answers to these questions, but I do believe an individual is the sum of all his or her actions—the sum of all his or her identities—and that knowing identity strategies for particular situations, even virtual ones, that work allows the individual to have a set of roles he or she can try out and redefine in new situations. If learning about the community and the self is tied into being able to try out different identities as a way to finding the optimal identity then I believe that a computer game allows trying on different roles or role-playing in a safer way than in many real-world communities. A game, then, would seem to be an ideal place for someone to learn what kind of identity works best (one who participates cooperatively) in a given community.

This idea of a community as a group of informed participants is similar to part of the third (of five) aspects which Colin Lankshear (1997) describes as defining New Capitalism:

The “vertically integrated large-scale organizations” of ‘old’ standardised mass production capitalism have given way to “vertical disintegration and horizontal networks between economic units.” This is partly a matter of . . . enlarged scope for workers to participate in decision-making (within definite parameters). It is also a matter of horizontal relationships of co-operation, consultation, co-ordination, in the interests of flexibility, decentralisation, and adaptability in production, which extend beyond the confines of a specific business or firm to include other ‘partners’ within an integrated productive enterprise . . .

This New Capitalist idea is essentially a demonstration of how people (and groups of people AKA commercial organizations) should participate in an economic Social Dilemma! Furthermore, for a New Capitalist economic model to work, the individual participants need to be more informed and better able to play different roles, in other words be more literate, than participants in previous Old Capitalist traditions.

---

9 In other words, the individual can shape-shift. See Gee (2002).
Games to Address Social Dilemmas

The difficulty in getting someone to “buy into” a particular way of thinking or to be a willing participant in a community is to get them to see him or herself as someone who wants to be and is a part of the community. Gee (2003) writes:

It has been argued that some poor urban African-American children and teenagers resist learning literacy in school because they see school-based literacy as “white,” as associated by people who disregard them and others like them. They don’t believe that a society that they view as racist will ever allow them to gain a good job, status, and power, even if they do succeed at school-based literacy. Thus they will not envisage themselves in the new identity that success in school-based literacy requires—that is, as the “kind of person” who learns, values, and uses such literacy and gets valued and respected for doing so. Without such an identity commitment, no deep learning can occur. The students will not invest the time, effort, and personally committed engagement that active, critical learning requires.

Gee (ibid) later suggests that the keys to allowing people to play with their identities and learn are:

1. The learner must be enticed to try, even if he or she already has good grounds to be afraid to try.
2. The learner must be enticed to put in lots of effort even if he or she begins with little motivation to do so.
3. The learner must achieve some meaningful success when he or she has expended this effort.

This applies to learning a particular literacy or learning to participate in and contribute to a community (which, as outlined previously, is really the same thing). Computer and video games already have the attraction needed to satisfy condition 1. Conditions 2 and 3 are easily addressed in a role-playing game by providing incentives to move the story along and to improve one’s character and then rewarding the effort it takes to follow-up
on those incentives. Additionally, the kinds of rewards given could be situated as community rewards to emphasize cooperation and participation in the community.

Many RPGs already have analogies to Social Dilemmas built into them, but I do not think developers are conscious of this and they sometimes include contradictory reward systems or present “good” and “evil” (i.e. cooperation and defection) in a heavy-handed fashion. What I propose is that educators and developers of CRPGs can consciously design in-game quests to focus on the idea of community involvement. Games of this nature would hopefully train players to be informed participants in a new social order.

References


