NOTES

1. Carl von Clausewitz, *On War* (Rockville, MD: Wildside Press, 2009), 24.

2. For more on the nature and morality of war, see chapter 14, "'Safe to Do What?': Morality and the War of All against All in the Arena."

3. David J. Lonsdale, "A View from Realism," in *Ethics, Law and Military Operations*, ed. David Whetham (Basingstoke, UK: Palgrave-Macmillan, 2011), 29.

4. Thucydides, *The History of the Peloponnesian War*, trans. Rex Warner (New York: Penguin, 1972), 402.

5. Brian Orend, *Morality of War* (Peterborough, NH: Broadview Press, 2006), 44.

6. For more on the sort of argument the Capitol might make to justify its harsh rule, see chapter 14, "'Safe to Do What?': Morality and the War of All against All in the Arena."

7. Anthony F. Lang, "Authority and the Problem of Non-State Actors" in *Ethics, Authority, and War*, eds. Eric A. Heinze and Brent J. Steele (New York: Palgrave-MacMillan, 2009), 52.

8. Michael Walzer, *Just and Unjust Wars*, 4th ed. (New York: Basic Books, 2006), 53–54.

THE TRIBUTE'S DILEMMA

The Hunger Games and Game Theory

Andrew Zimmerman Jones

In Panem, games are no child's play. At the center of Panem's society is the brutal and deadly annual event that gives the Hunger Games trilogy its name. It's a far cry from the games that most of us play in our own society. Or is it?

In fact, for nearly a century, mathematicians have been examining many aspects of our society, from economic systems to international relations and warfare, by treating them as games. The analysis of the rules of various games and how to formulate effective strategies is called *game theory*. The science of game theory can teach us about the world around us, and it can also provide us with key insights into the motivations and actions as well as the alliances and betrayals played out within the Hunger Games trilogy.

The Game with the Bread

In game theory, a game is defined as any situation in which two (or more) decision makers (called *players*) face off against each other, resulting in a change in their status. Game theory works by breaking the game's rules into a series of allowed strategies, with the analysis providing a guide for which strategy is the best one in a given circumstance. Mathematical models track the various payouts and losses of different strategies. The Hunger Games are structured so that even the most favorable outcomes are bad, but the better you can analyze and quantify the possible strategies, the better your chances are of surviving.

The simplest type of game is called a *zero-sum* game, in which a victory for one player is an equal loss for the other player. Exploring this simple game gives us an opportunity to see the mathematics behind game theory. We can imagine a young Katniss Everdeen and a young Peeta Mellark playing a zero-sum game in which they each hold up one or two fingers. If the sum of their fingers is even, Peeta gives Katniss a piece of bread. If the sum is odd, Peeta wins a piece of bread from Katniss. This is actually a classic children's game called Two-Finger Morra. It's useful to depict the possible wins, losses, and strategies in the form of a table, called *strategic form*, as shown in table 1.

The numbers in table 1 represent the bread gained or lost by each participant. The first number in each set of parentheses represents Katniss's outcome, and the second number is Peeta's outcome. We see that if both hold up one finger, the sum is two, an even number, which means Peeta loses a piece

Table 1. Two-Finger Morra

If the sum of the fingers is even, Katniss wins a piece of bread from Peeta. If the sum of the fingers is odd, Peeta wins a piece of bread from Katniss.

		Player 2: Peeta	
		Holds up 1 finger	**Holds up 2 fingers**
Player 1: Katniss	**Holds up 1 finger**	(1, −1)	(−1, 1)
	Holds up 2 fingers	(−1, 1)	(1, −1)

of bread to Katniss. This is the value (1, -1) in the top left cell of the table: Katniss gains (1) and Peeta loses (-1). The profits and losses within each cell cancel each other out, adding up to zero, which is where the term *zero-sum* comes from.

An important aspect of this game is that neither player has any special advantage. Only by sheer luck will anyone get significantly ahead playing multiple rounds of the game. There's no strategy that either Katniss or Peeta could apply to improve her or his chances of winning.

Now consider the same game, but instead of holding up one or two fingers, each player can now hold up as many as three, as shown in table 2.

This scenario is very different from the first one, because strategies can now be applied in an effort to maximize each player's chance of success. Katniss will most likely realize that when she holds up two fingers, she has two chances of losing bread but only one chance of winning. By holding up one or three fingers, however, she reverses those odds. The problem with this plan is that it assumes that Peeta isn't going to

Table 2. Three-Finger Morra

This game has the same rules as Two-Finger Morra, but now each player can hold up as many as three fingers. Unlike in the Two-Finger version, the odds are now in Katniss's favor, since there are more even combinations than odd.

		Player 2: Peeta		
		Holds up 1 finger	**Holds up 2 fingers**	**Holds up 3 fingers**
Player 1: Katniss	**Holds up 1 finger**	(1, −1)	(−1, 1)	(1, −1)
	Holds up 2 fingers	(−1, 1)	(1, −1)	(−1, 1)
	Holds up 3 fingers	(1, −1)	(−1, 1)	(1, −1)

figure out that when he puts up one or three fingers, he has two chances of *losing* bread. Peeta isn't just a passive part of the environment; he's an independent agent. One of the key aspects of game theory—and the thing that makes it so interesting—is that "while decision makers are trying to manipulate their environment, their environment is trying to manipulate them."[1]

Granted, Peeta might throw the game because he's smitten with Katniss. But assuming he's playing to win, his most reasonable strategy is to put up two fingers, because that will be a winning strategy as long as Katniss sticks to one or three fingers. Before long, though, Katniss will figure out what Peeta is doing and begin raising two fingers. When that happens, Peeta will realize that he needs to mix things up, and eventually they'll both arrive at the strategy of randomly choosing among the three options. Still, even if Peeta played his best strategy, the odds are ever in Katniss's favor in this variant of Three-Finger Morra, because there are more combinations that result in even results than odd ones.

The two games we just discussed are fairly easy to analyze, because we've defined the payoff and the loss in a way that can be easily quantified or assigned numerical values. Most of the games within the world of the Hunger Games trilogy are much more complex, however, partly because there are more than two players, but also because the payouts include harder-to-quantify rewards like sponsorships and the prospect of living a few hours longer.

The First Game: Reaping by the Numbers

Every aspect of life in Panem seems to revolve around the Hunger Games. Even citizens who are not participating in the Games directly are caught up in a complex web of incentives and penalties all tied to the Hunger Games. One strategic decision that the citizens of Panem have to make is whether to increase their odds of being reaped in exchange for food.

Previously, we defined a *game* as having two decision makers, but in this case the only decision maker is the individual. The second player has no decision-making power. As at a Las Vegas blackjack table, the rules of the establishment are set in stone by the house. Let's start by analyzing the actual odds that Katniss will be reaped. District 12 has about eight thousand residents. In line with Katniss's observation that "almost no one can afford doctors," there's no indication of a large elderly population.[2] Let's make the generous assumption that most District 12 residents are lucky to make it to sixty. To simplify, let's equally divide the residents by age, giving us about 134 residents of each age within District 12. Since young people are eligible for seven years (from ages twelve through eighteen), in any given reaping, there would be about 938 kids in the reaping pool.

Most of the young people will be entered more than once, though, since each name goes into the pool an additional time for each subsequent year of eligibility. Also, poor youths can choose more entries in exchange for tesserae, each "worth a meager year's supply of grain and oil for one person."[3] At the beginning of the first book, Katniss's name is in the drawing twenty times (three tesserae for five years) and Gale's a whopping forty-two times (five tesserae for seven years). Although there are undoubtedly many orphans in District 12, we can assume that most of the youths in the reaping will still have two living parents (or else our estimate of the upper age is way off). That means that they will probably require an average of about three tesserae a year. Those who don't need tesserae, like Madge and Peeta, will be balanced out by those, like Gale, who have larger families.

Running through the numbers and remembering that entries are cumulative from year to year, we get about fifteen thousand entries in each reaping.[4] Based on this estimate, the probability of Katniss being chosen is 0.13 percent (thirteen in ten thousand), with Gale's probability at 0.28 percent (twenty-eight in ten thousand). Madge and Peeta, who don't have to trade for tesserae, have a probability of only 0.03 percent (three in ten thousand). Katniss's chances of being chosen for the Hunger Games more than quadruple if she takes tesserae for her family.

We don't need to build a table to see that these are very small percentages, all less than 1 percent. If *not* taking the tesserae will increase the likelihood that any given family member will die by as little as 10 percent, then it's easy to understand why someone would make such a devil's bargain. Taking the tesserae produces a net increase in one's odds of survival. Although taking the tesserae drastically increases the chances of being reaped (a ninefold increase for Gale), those chances are small compared to the chances of starving without the tesserae.

Of course, no matter how small the odds, you're always gambling with your life, and the most you can ever hope to win is the prolongation of an impoverished and oppressed existence. In fact, even the Hunger Games victors don't get a net positive outcome, because they spend their lives luxuriously enslaved to the Capitol. This game is certainly not zero-sum. As at a Las Vegas blackjack table, the odds *always* favor the house—or, in this case, the Capitol.

Consider also what the first game shows us about Katniss. Despite all efforts by the Capitol, there are still those like Katniss who refuse to play by the strict rules the Capitol has set up. When we first meet Katniss, she is heading off into the woods, where she hunts to supplement her meager food supply. Katniss constantly looks for ways to modify (or break) the rules in her favor so that she has a better chance of getting out alive.

The Second Game: Training Days

Once the tributes have been reaped, they travel to the Capitol for the Training Days, where they not only train but also show off their talents so that mentors can line up sponsors to provide aid during the Games. The mentors' role is crucial here: they are the intermediaries between the tributes and the sponsors.

As we learn, though, mentors also have a difficult choice to make, because they can't play up both tributes equally when promoting them to the sponsors. There can ordinarily be only one survivor, and the mentor has to choose which tribute to support fully if he or she is going to have any chance of getting someone out alive.

One of the defining aspects of a game is that the player's strategy influences the environment, which in turn influences the player's choices of strategy. In making an appeal to the sponsors, the players try to alter the game landscape in their favor. The Training Days represent what game theorists call a *subgame*, "a well-defined game within a game."[5] The outcome of the subgame will help to shape one's odds in the final, fatal game that takes place in the arena.

In the grand scheme of the Hunger Games, the tributes probably have more control over the outcome of the Training Days subgame than the actual arena battle, which is dictated largely by the whims of the Gamemakers. During the Training Days, the tributes can plan their costumes, garner support through the interviews, and choose which skills to demonstrate and which to keep under wraps. The Training Days give each tribute the opportunity to shine in his or her own way.

Again we see that one of the key elements of this game is that the Capitol, in the form of the Gamemakers, is the final arbiter of each tribute's score. The Gamemakers are the ones who quantify each tribute's performance. Even in this tribute-driven subgame, the Capitol is able to exert its influence with an eye toward the outcome it wants.

The Prisoner's Dilemma: Cooperate or Betray?

The Training Days are also an opportunity to meet the other tributes, learn more about them, and possibly form strategic alliances. These alliances are always tenuous, betrayal being inevitable unless one party dies at someone else's hands during the game. Otherwise, your partner is bound to betray you unless you betray him or her first.

This situation parallels a classic game-theory problem known as the prisoner's dilemma, which is presented in table 3 with a Panem twist. Consider two suspected rebels captured by the Capitol Peacekeepers. If neither rebel confesses, they will only be charged with a minor offense and serve a short jail sentence, such as one year. However, the Peacekeepers offer each rebel a deal. If one rats the other out, the Peacekeepers will drop all charges against the rat, but the other rebel will get ten years in prison. If they both rat each other out, both get charged with the crime and end up with about eight years in jail. This scenario is shown in table 3.

The payoff of zero years is called the *temptation* in this game, and the ten-year sentence is called the *sucker's payoff*.

Each rebel's best strategy is to betray his or her partner.[6] A rebel who betrays while his or her partner cooperates gets no prison time instead of one year. If the partner also betrays, each gets only eight years instead of ten. The betray-betray outcome is therefore what game theorists call an *equilibrium point*, defined as "a stable outcome of a game associated with a pair of strategies. It is considered stable because a player unilaterally picking a new strategy is hurt by the change."[7] In other

Table 3. The Prisoner's Dilemma

The best option for each player is to betray the other, but both players come out ahead if they cooperate with each other in remaining silent.

		Rebel 2	
		Cooperate	Betray
Rebel 1	Cooperate	(1, 1)	(10, 0)
	Betray	(0, 10)	(8, 8)

words, betrayal always gets you less jail time than cooperating, regardless of what the other rebel does, so one's best option is always to betray—or so it would seem!

The problem is that the equilibrium point, betray-betray, doesn't provide the best overall outcome. It gives each prisoner eight years, whereas a much better outcome of only one year apiece is available if they both cooperate. If only they could trust each other and coordinate a cooperative strategy, they could get a better payout. That plan is dangerous, though, because the incentive to betray is so powerful. Suppose rebel 1 has absolute trust that rebel 2 isn't going to betray. What's to prevent rebel 1 from unilaterally switching to a betrayal strategy to avoid having to serve any jail time at all? Rebel 2 is now a sucker doing ten years in prison while rebel 1 is a clever betrayer. If they're equally clever, though, they'll both betray and thus do eight years apiece in a Capitol prison, assuming that the Peacekeepers hold to their deal, of course.

The point of this classic problem is that sometimes the strategy of everyone looking out for himself or herself leaves everyone worse off than if they'd worked together. But the most interesting aspect of the prisoner's dilemma game comes to light when we devise a strategy for playing it over and over, taking into account our past experience with the other player. "It is when the prisoner's dilemma is played repeatedly," according to game theorist Morton Davis, "that the cooperative strategy comes into its own."[8]

Game theorists have created computer simulations to test various strategies for repeated plays of the prisoner's dilemma. These simulations are designed so that each player follows a certain programmed code that uses its opponent's past behavior to determine which strategy—cooperate or betray—to implement in a given round of the game. The prisoner's dilemma plays out repeatedly, with the cumulative amount of time that the player spends in prison tallied up to measure how well a given strategy works.

There are three obvious strategies: always cooperate, always betray, or randomly pick (about 50 percent of each). But computer simulations have shown that none of these strategies plays out best in the long run. The most successful strategy is the simple but elegant tit for tat.

The tit-for-tat strategy starts out with cooperation but then mimics the last strategy employed by the other player. If the last strategy was betrayal, then tit for tat will use betrayal in the next round. If the other player cooperated in the last round, tit for tat will use cooperation in the next round. If a rebel using the tit-for-tat strategy is betrayed 100 times, but the other player then cooperates on the 101st round of the game, the rebel will cooperate on the 102nd round. The 100 previous betrayals won't matter, but neither would 100 previous cooperations.

Tit for tat usually ties with a rival strategy or even loses by a few years. For example, if played against the always-cooperate strategy, tit for tat would never betray, and the two strategies would perfectly tie. Paired off against the always-betray strategy, tit for tat would cooperate once, getting the sucker's payoff, and then betray for the rest of the game, losing in the final tally because of its loss in the first round. It would on average tie with the randomly-pick strategy.

How, then, can tit for tat be called the best strategy? It's because of the way victory is calculated in these simulations. Each strategy is tested against all of the others many times, and the total time in prison is counted up. Against a broad range of strategies, the tit-for-tat rebel spends the least amount of time in the Panem prisons, because this strategy doesn't have an inherent flaw that can be exploited by the others, particularly by always betray. Always cooperate, for example, is profoundly vulnerable to always betray, but always betray is also vulnerable when played against itself. Randomly pick can't gain from knowledge of past behavior and so is also vulnerable to always betray. Consequently, these strategies often end up accumulating much prison time.

If you play hundreds of simulated games, pairing diverse strategies against each other, tit for tat will generally spend

the least time in prison. But there's one problem employing tit for tat in a real-world prisoner's dilemma scenario: the players don't always know whether they're being cooperated with or betrayed, as evidenced by Katniss's complete misunderstanding of her allies throughout the series.

The Arena: The Tribute's Dilemma

To explore how this applies to the Hunger Games, let's modify the prisoner's dilemma into the tribute's dilemma. By forming an alliance (cooperating), the tributes gain a temporary benefit: an increase in their overall prospects for survival. But this cooperation strategy can't continue indefinitely. The rules dictate that if both survive long enough, at some point a betrayal strategy must be implemented by one of the participants. So at each moment of an alliance, the players are looking for safe ways to betray their allies while also guarding against being betrayed.

But what about tit for tat? Applied to the arena, it would require cooperating until your ally betrays you, either by trying to kill you or by not protecting you from an attack. Wouldn't this strategy work? Unfortunately it wouldn't, even if you have better judgment than Katniss about who your allies and enemies really are. (Recall that Katniss spends much of her time in the arena—during both games, in fact—believing that she has been or will be betrayed by people who are actually among her most loyal allies.) The problem is that in most cases, if someone betrays you in the arena, you won't live long enough to carry out the tit for the betrayer's tat.

In their first Hunger Games, Katniss and Peeta are given a glimmer of hope that an always-cooperate strategy might work for them when it's announced that there may be two victors. But that hope is denied them in the final moments of the game. When Katniss hears again that only one will be allowed to exit the arena alive, she reflexively pulls her bow on Peeta, once again displaying her lack of understanding of his motivations. Peeta, being in love with Katniss, is absolutely unable to implement the betrayal strategy. But it turns out that Katniss can't implement it either, both because it's against her nature to murder an unarmed person in cold blood and because she has developed genuine feelings for Peeta.

Let's return to an observation we made earlier when we were talking about starving families in District 12: Katniss is the sort of person who looks for ways around the rules. The rules of the game allow only a strategy of betrayal, but she realizes that full cooperation is still possible by going outside the parameters of the game that the Gamemakers want her to play. Look at it this way: Katniss would surely see the flaw in the underlying assumptions of the classic prisoner's dilemma situation. The game is defined so that there are only two players, the two rebels pitted against each other, but if Katniss were interrogated, she would realize that there's a third player involved, the real enemy: the Peacekeepers. Or, in the case of the tribute's dilemma, the Gamemakers, the Capitol, and the very Games themselves.

In the final moments of *The Hunger Games*, Katniss perceives that she's in a position to deny the Capitol victory by embracing mutual cooperation to the point of suicide, stepping completely outside the rules of the game. She's willing to gamble that the Gamemakers can't afford to not have a victor—or, in game theory terminology, to have a payout of zero: "We both know they have to have a victor. Yes, they have to have a victor. Without a victor, the whole thing would blow up in the Gamemakers' faces. They'd have failed the Capitol. Might possibly even be executed, slowly and painfully, while the cameras broadcast it to every screen in the country: If Peeta and I were both to die, or they thought we were . . ."[9]

By refusing to play by the established rules, Katniss forces all of the players into a new game, the ramifications of which reverberate through the second and third books. This reorientation of players uniting against a higher-order opponent is a regular theme throughout the series, although it's not always

Katniss who implements this strategy. In *Catching Fire*, there's an elaborate alliance behind the scenes, the scope of which Katniss is entirely unaware of, and again the key to victory lies in changing the rules of the game, this time by busting the tributes out of the arena at the last moment. The climax of *Mockingjay* finds Katniss standing firm against an enemy that only she (and maybe Haymitch Abernathy) fully recognizes: President Coin. Betrayed by Coin, Katniss decides to implement a little tit for tat and betrays Coin in return, terminally.

Each time that Katniss is put in a situation in which her options seem to point at a net loss, she creates a whole new strategy that changes the very environment of the game and tilts things in the direction she wants. She's reminiscent of other famous innovators who stepped outside the established paradigm and took off in new and unexpected directions. Henry Ford (1863–1947) once said that if he'd given the people what they wanted, he'd have made a faster horse. The Nobel Laureate physiologist Albert Szen-Gyorgi (1893–1986) captured this idea when he said, "Discovery consists of seeing what everybody has seen, and thinking what nobody has thought."[10] The mathematician John Nash did this for game theory when he realized that sometimes the optimal solution doesn't yield the best results, as in the prisoner's dilemma. The brilliance of Katniss consists of that same innovative spirit, her ability to step outside the established logic and view some aspect of reality from a new perspective.

The world described by mathematical models in game theory isn't nearly as interesting (or dangerous) as the world we enter when those models break down. Although we can model the world with game theory, there's no requirement that the world follow those rules, any more than the characters in the Hunger Games trilogy must play by President Snow's rules. The rigid rules of game theory do not apply with certainty to either life or the Hunger Games. From a game-theory standpoint, that makes them very poorly defined games, but it makes for an interesting world and a great trilogy of novels.

NOTES

1. Morton D. Davis, *Game Theory: A Nontechnical Introduction* (Mineola, NY: Dover, 1983), 61.
2. Ibid., 8.
3. Ibid., 13.
4. Here's how I arrived at this number: 4 entries per youth times 134, then multiply that times the sum of how many years all of the participants have been entered (1 + 2 + 3 + 4 + 5 + 6 + 7 = 28); that yields 15,008. It'll actually be a bit less, because each year the entries for two people are taken out when they get reaped.
5. Edward C. Rosenthal, *The Complete Idiot's Guide to Game Theory* (Indianapolis, IN: Alpha Books, 2011), 63.
6. This strategy is often called *defect* in the literature, but I think *betray* fits better with the Hunger Games.
7. Davis, *Game Theory*, 14.
8. Ibid., 112–113.
9. Suzanne Collins, *The Hunger Games* (New York: Scholastic Press, 2009), 344.
10. Quoted in Irving John Good, ed., *The Scientist Speculates: An Anthology of Partly-Baked Ideas* (New York: Basic Books, 1962), 15.

PART SEVEN

"IT MUST BE VERY FRAGILE IF A HANDFUL OF BERRIES CAN BRING IT DOWN": THE POLITICAL PHILOSOPHY OF CORIOLANUS SNOW