Runaway Train Carney Game
An Amoral Prisoner's Dilemma Narrative

By Greg London
http://www.greglondon.com
Rev: 28 July 2013

There's a game theory concept called a "Prisoner's Dilemma" that is described like this:

> Two members of a criminal gang are arrested and imprisoned. Each prisoner is in solitary confinement with no means of speaking to or exchanging messages with the other. The police admit they don't have enough evidence to convict the pair on the principal charge. They plan to sentence both to a year in prison on a lesser charge. Simultaneously, the police offer each prisoner a Faustian bargain. If he testifies against his partner, he will go free while the partner will get three years in prison on the main charge. Oh, yes, there is a catch ... If both prisoners testify against each other, both will be sentenced to two years in jail.

The payoff matrix is generally represented like something shown below. The numbers in the matrix indicate the number of years each prisoner will end up in prison.

		Player B	
		betray	cooperate
Player A	betray	2 2	3 0
	cooperate	0 3	1 1

The point of the Prisoner's Dilemma is to present a scenario in mathematical terms where the individual actors acting in their own self interest will end up at an outcome worse than if they both cooperate.

That's the *point* of the Prisoner's Dilemma.

The problem is that it is extremely difficult to get some people to understand this. Some people will look at the Prisoner's Dilemma scenario and will mistakenly think that individuals acting in their own self interest will cooperate, and will both accept 1 year in prison rather than try to betray the other prisoner.

After having lots of conversations about the Prisoner's Dilemma with many different people over many years, I think one of the fundamental hurdles with people understanding the Prisoner's Dilemma is that the when a normally-law-abiding person imagines themselves trapped in the Prisoner's Dilemma scenario, they see it as, fundamentally speaking, immoral and evil. For police to force an innocent person into the Prisoner's Dilemma scenario, they have to be crooked cops, because the guilt/innocence of the prisoner's becomes irrelevant to the

outcome. If someone who generally follows a moral world view is asked to imagine themselves in the Prisoner's Dilemma scenario, they may approach the scenario with the notion that they are innocent, they have committed no crime. And as a result they choose to claim their innocence, because that is the moral, right, and proper thing to do.

When a normally law-abidng, innocent person imagines themselves trapped in the Prisoner's Dilemma, the underlying narrative of the Prisoner's Dilemma becomes an immoral tale. Innocent people are threatened with prison by crooked cops who don't care if the accused are innocent or guilty but only care about throwing people in jail for as long as possible.

I believe this underlying immoral narrative triggers in quite a few people an urge to resist and oppose the immoral scenario itself, rather than engage in the game theory simply based on cost/payoff matrix. In other words, I think quite a few people choose to claim their innocence and to not betray the other innocent prisoner in an effort to stand up to the immorality of the crooked cops.

This partly occurred to me while I was watching "The Dark Knight" in 2008. In the movie the Joker (played by Heath Ledger) loads up two ferry boats with explosives, gives each boat a detonator for the *other* boat, and then tells everyone on both boats that if one boat doesn't blow the other boat up by midnight, the Joker would blow both of them up. The Joker is chaotic evil. And the people on both ferries are presented as initially wanting to blow up the other ferry to save themselves, but in the end, they choose instead to stand up to the Joker's evil plan, even if it meant they all end up dying to stand up to evil.

It was the moral thing to do. But the problem is, studies have shown that people making choices based on economic considerations (what's best for me), may make entirely different choices if that same choice is presented with a moral backdrop (what's best for everyone overall). If you were trapped on that boat and viewed your choices from the "What's best for me" calculus, you would blow up the other boat. But doing that would result in killing people, which invokes the moral calculus to do the right thing, so, neither boat blows up the other.

There are real-world examples of some situation being viewed from the economic view and the moral view, and the different outcomes it produces. In 1993, Switzerland asked a small mountain village named Wolfenschiessen to support putting a nuclear waste repository nearby. About 51% of the people supported it. Trying to increase support, Switzerland offered all residents a small cash payment if they would accept. Support DROPPED to 25%. The government had initially invoked a moral decision for the residents and residents chose from a “What’s best for everyone” calculus. Everyone would be better off if someone accepted having a nuclear waster repository in their backyard. When the government added a monetary incentive to increase support, they unwittingly changed the decision making calculus of the residents from “what’s best for everyone” to “what’s best for me”. And suddenly, having a repository in your backyard didn’t seem worth the small amount of money they were offering.

http://www.bostonreview.net/forum-sandel-markets-morals

People's decision-making process can be radically different depending on whether they're making a moral decision or economic decision.

And I believe that one hurdle that prevents some people from understanding the Prisoner's Dilemma is that when they imagine themselves trapped in that situation, they see it as a purely moral choice. If they're innocent, then the entire scenario is evil and they choose to

do the moral thing and stand up to it and resist it. And given that their choices may put someone else in jail for a long time, someone they may view as innocent like themselves, that also invokes a moral decision-making process. And because they see it from a moral perspective, they think everyone will naturally cooperate and achieve the best outcome for both players. But since they see it from a moral perspective, they don't see it from the economic perspective; they don't see that if both players choose "what's best for me", that they choose to betray the other prisoner, ending up in the worst possible outcome for all.

So, I wanted to create a game theory scenario with the same payoff matrix of the prisoner's dilemma scenario but that involved a narrative that wasn't itself immoral, that was instead amoral. The cost/payoff matrix would be the same, but the story used to setup the scenario and create the cost/payoff matrix wouldn't be based on some immoral setting.

Without an immoral narrative setting up the decision process, hopefully people would be more likely to approach the scenario from an economic standpoint, what's in their best personal interest. And then they would see how people operating in their own best self interest can end up creating a much worse outcome than if they cooperated.

The main thing I wanted to do was change the narrative so that the *cause* of the players having to choose betray/cooperate was not due to evil third-party people. In the Prisoner's Dilemma, players are forced into choosing to cooperate or betray the other prisoner because crooked cops pick them up and haul them into the station. I wanted some kind of scenario that wasn't invoked by evil third party individuals, but rather was caused by some thing that didn't think. For example, gravity isn't evil, gravity is amoral, and by making the cause of the scenario something physical, something that was a force of nature, this could go a long way to avoiding invoking a moral decision in people, and therefore allow them to approach it from an economical standpoint.

The other thing I wanted to change was the labels of the choice itself. In the Prisoner's Dilemma, each player is presented with a choice: either they "cooperate" with the other prisoner, or they "betray" the other prisoner. Both terms have quite a bit of moral baggage attached to them. "Betray" has bad connotations attached to it. "Cooperate" has good connotations attached to it. Both may invoke a moral decision, making it difficult for people to see the economic dilemma that the Prisoner's Dilemma is trying to present. So I wanted the choice to be between two terms that were amoral as well.

Basically, what is needed in the physical scenario are two interlocks so that player 1 affects player 2 and player 2 affects player 1. When you have something like crooked police creating the scenario, the interlock is the crooked court system, so the interconnections aren't quite as obvious. In physical scenarios, these interlocks have to be implemented using multi-way switches, multi-way valves, gears, linkages, or similar. So, I tried to come up with a simple solution based on a hypothetical carnival game.

So, here is my game called the Runaway Train Carney Game:

Runaway Train Carney Game

The carnival is in town and you've decided to purchase a $50 admission ticket and check out the carnival. While you're walking around, a vendor waves you over and says that you have been randomly selected for one free play at the "Runaway Train" game.

Just an aside, some Carney games are rigged. This game is not rigged.

The vendor hands you a ball and tells you that's the "train". Then he shows you a wooden board with grooves cut into it and he explains that the grooves are the "tracks".

The game is a two player game. The other player also gets a ball ("train") and they have their own wooden board with grooves ("track") that is laid out exactly like yours.

From the starting position, the ball will roll down to a fork in the track. Each player gets to select whether their ball will go left or right. If the player selects right, their ball will go to a second fork that selects between a prize of $30 or $10. If the player selects left, their ball will go to a third fork that selects between a prize of $20 or nothing. You do not get to control the second and third forks of your track. The other player controls them.

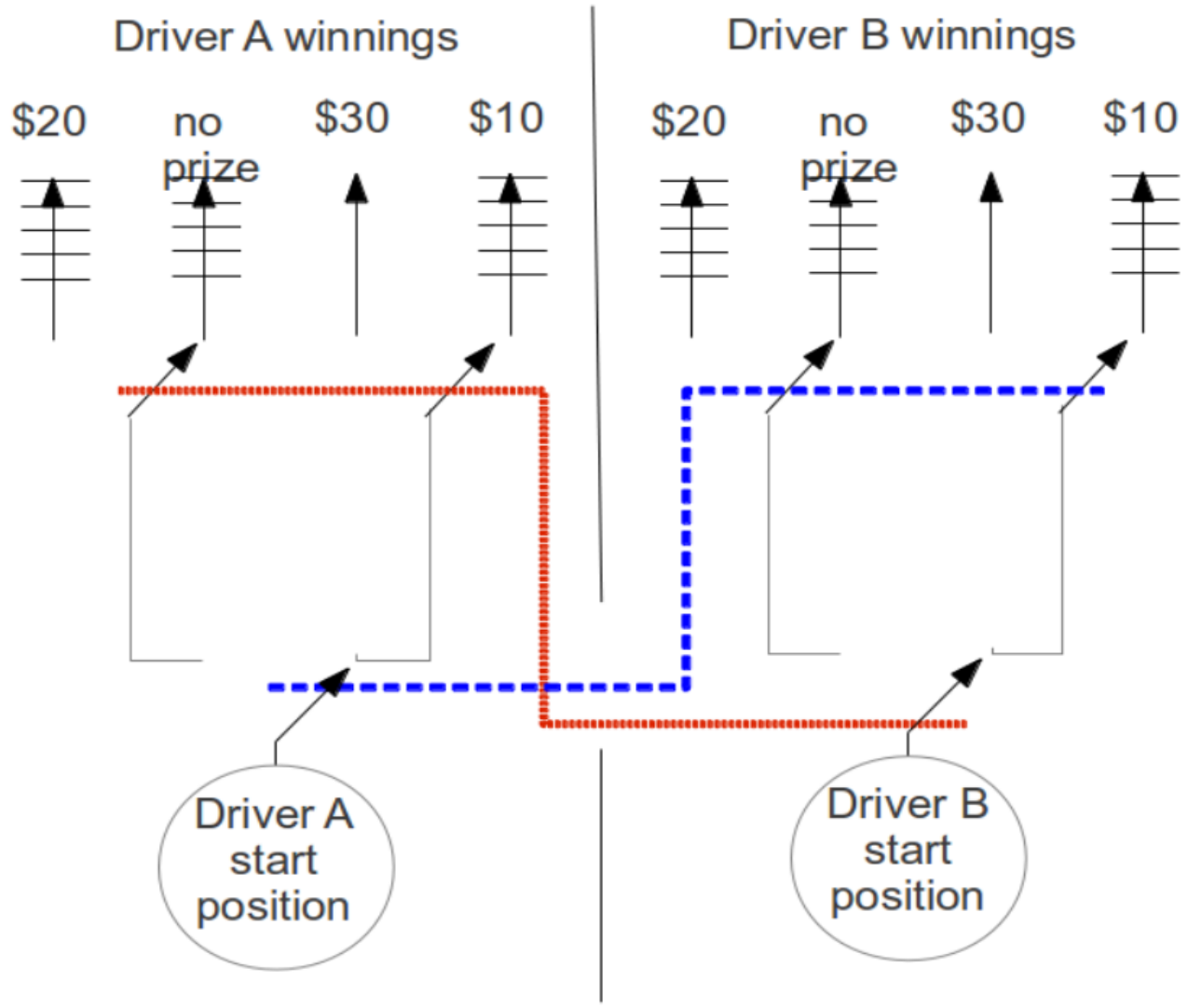
The second player has exactly the same track and scenario. They get to select the first fork going left or right. And the second and third forks are controlled by whatever you choose for your first fork.

When you are ready to play, the vendor drops down a panel which covers up the second and third forks so you can't see what the other player is choosing. The vendor tells you to put the ball in the ball in the starting position, then move the lever to select left or right. When you and the other player have both made your selections, the balls start rolling, the selections are locked into place, and the panels lift up to reveal the other player's selection.

Once you've played, they punch a hole in your ticket and you can't play again without buying another $50 admission ticket.

One last thing. The game is set up in a long, long structure. The other player is at the other end of the structure. And other vendor games and booths extend from both sides of the structure for many, many booths (several hundred feet). The carnival is packed, so you can't see or hear the other player. Since the other player is also randomly selected, you have no idea of knowing who the other player is. And once the game is over, it will take you several minutes to get around to the other side to see who the other player was. And at that point, they will probably have walked off.

The track layout looks like this:



The start positions are physically higher than the end of the tracks where the prize is selected.

The payoff matrix looks like this:

		Player B	
		left	right
Player A	left	$20 $20	$30 no prize
	right	no prize $30	$10 $10

The Amoral of the Story:

I believe this scenario presents the payoff matrix that is inherent in a Prisoner's Dilemma but presents it in a purely economical narrative and should invoke a purely economic decision-making process.

If the other player is completely outside of your control, if you can't communicate, threaten, bribe, or retaliate against the other player, if you can only play the game once, then your choice is right($30/$10) or left($20/$0), and you should choose right. The same choice is presented to the second player, and they should also choose right ($30/$10). And the result is you both get $10.

If you both chose left, you'd both get $20. But if either one of you chooses left, the other player could choose right and get even more money ($30), leaving the other player with no prize at all. This is what makes it unstable. If you try to work together to get $20, the other player has incentive to betray you and get even more money. So, both players choose right($30/$10) and both end up with $10.

To use economic terms, since you don't know what the other player will do, and they might take advantage of you choosing left, you have to "hedge your bet" and choose right. You'll only get $10, but the alternative could leave you with no prize at all.

This is the inherent instability of the Prisoner's Dilemma. If you both "cooperate", then you'd both be better off than if you both "betray". But if one player "cooperates", then the other player has even more incentive to "betray". This creates the unstable outcome. So, you both "betray", and you're both worse off than if you "cooperate".

The problem in understanding this in the Prisoner's Dilemma, I believe, is that "cooperate" and "betray", and the inherent immoral narrative of the Prisoner's Dilemma (dirty cops trying to convict innocent people), causes quite a few people to make a moral choice and "cooperate". But when the narrative is changed to an amoral explanation, when the choices are amoral like "left" and "right", then all that's left is an economic decision.

Hidden Complexity Made Plain:

I think this scenario also highlights something that may not be quite so obvious from the Prisoner's Dilemma scenario: this game is rather complex. You are given a choice of left or right which controls the first fork of your "train". Right will give you $30 or $10. Left will give you $20 or no prize at all. But the board in front of you has three forks, two of which are out of your control. And there is another player who only controls one fork for his "train", and the other two are controlled by you.

From an individual player's point of view, the only thing a player controls that affects their winnings is one fork. But the game consists of a total of six forks and a second player who is completely out of the other player's control.

When presented using the Prisoner's Dilemma narrative, the choices, the six different forks, the complexity of the track, gets replaced with dirty cops. The "interlock" between the two players is explicit in the "runaway train Carney game". The interlock between players in the "Prisoner's Dilemma" is merely implicit in the narrative, and is left for the players to intuit.
Partner's in Crime:

Sum-body Loves You:

So, I presented an earlier version of this scenario to someone, and the first question they asked was "Who is the other player?" I originally thought this might indicate they were still making a moral decision. But their only concern was whether or not the other player was their spouse/significant other. The thought process is something along the lines of "what's mine is theirs. What's their's is mine." If the other player was a significant partner, then the payoff matrix no longer cares about the individual prize amounts, but rather the SUMS.

From the point of view of the "Prisoner's Dilemma", I didn't look at the sums of the prison sentences. In hindsight, it makes sense to consider it this way. But from the point of view of a carnival game with prize money, it is actually almost natural to imagine two partners playing the game, looking at the sum of money they get in total, and realizing that they should both pick left.

Here is the payoff matrix with sums included in **bold**:

		Player B	
		left	right
Player A	left	$20 **$40** $20	$30 **$30** no prize
	right	no prize **$30** $30	$10 **$20** $10

If the other player is someone you share money with, then the payoff matrix is no longer an unstable scenario. Both players will choose "left" because that will win the couple the most money as a couple. Both players will get $20, and the couple will get the maximum payoff of $40. Any other combination of choices results in a lower sum than $40.

By looking at the sums, the game immediately changes from unstable, to completely stable. If you know your money-sharing partner is the other player in your game (and if they know you're the player in their game), then you know you should pick left to maximize your money as a partnership.

I think looking at the sums is a way to look at how the moral calculus works. The moral calculus is "what's best for everyone". From that perspective, the focus isn't on what I get, but what all the players get as a whole, i.e. the sum of the prizes. This moral calculus can occur to the point that an individual player may be willing to accept an outcome that is less-than-ideal for them personally if it means that all the players overall get a better deal.

This was perfectly demonstrated by the town of Wolfenschiessen being willing to accept a less-than-ideal individual outcome of having a nuclear waste repository put near their town because from the moral calculus, everyone as a whole was better off.

If the moral perspective is looking at the sums, it's not sufficient to say "The needs of the many outweigh the needs of the few". Rather, if the benefit to the many is large enough, a few individuals may actively choose a small sacrifice for themselves.

Summary:

My goal was to come up with a game that creates the same payoff matrix as the Prisoner's Dilemma, but explains itself with a completely amoral narrative to set it up. I believe the Runaway Train Carney Game does just that.

The reason why I wanted to create this amoral game was to make the complexity of the game explicit (mechanically visible) and to make it easier for people to see the economic decision and to get people to avoid defaulting to the moral decision-making process.

From an economic decision-making process, the Prisoner's Dilemma and the Runaway Train Carney Game show a situation where two people acting in their own self interest end up choosing paths that create the worse outcome for everyone.

I believe the Runaway Train Carney Game makes the economic decision-making process easier for some people to see and therefore makes it easier for people to see the conundrum inherent in the scenario.