Interlocked Marble Race

The Mechanical Prisoner's Dilemma Puzzle

By Greg London

www.greglondon.com/imr

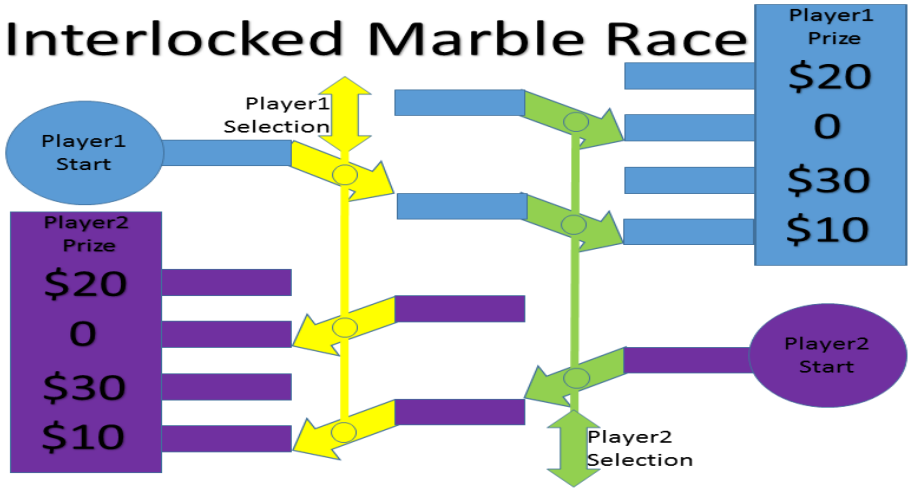
Current Revision: 19 December 2015

First rev: 11 December 2015

THE GAME: The carnival is in town and you've decided to purchase a $50 admission ticket and check out the scene. While you're walking around, a vendor waves you over and says that you have been randomly selected for one free play at the Interlocked Marble Race game. They ask to see your entrance ticket, and then punch a hole in it and tell you that you can only play once.

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

The vendor hands you a marble. And shows you a wooden board with grooves cut into it. There is a handle that you get to choose “up” or “down”. On the other side of the trailer, on the other side of a wall so you can’t see, another randomly selected player is given the same setup. The carneyman explains that the board operates like this:



Player 1 will put their marble in the point marked “Player1 Start”, and the marble will then go through a downward sloping blue track, through a yellow switch, and then through one of the two green switches, until the marble ends at one of the prizes in the blue box labeled Player1 Prize.

Player 2 will put their marble in the point marked “Player2 Start”, and the marble will then roll through a downward sloping purple track, through a green switch, and then through one of two yellow switches, until the marble ends at one of the prizes in the purple box labeled Player2 Prize.

Player1 has a selector for “up” and “down, represented by the yellow, two-ended arrow marked “Player1 Selection”. This selection controls 3 switches in the marble tracks marked in yellow. 1 switch affects the path Player1’s marble takes. The other 2 switches only affect Player2’s marble. If Player1 selects “up”, all 3 yellow switches go up. If Player1 selects “down”, all 3 yellow switches go down.

Player2 also has a selector for “up” and “down, represented by the green, two-ended arrow marked “Player2 Selection”. This selection controls 3 switches in the marble tracks marked in green. 1 switch affects the path Player2’s marble takes. The other 2 switches only affects Player1’s marble. If Player2 selects “up”, all 3 green switches go up. If player2 selects “down”, all 3 green switches go down.

There is a panel that separates both Players, so they can never see each other and never know who the other player is. The players are at opposite ends of a carney trailer, and the trailer is in the middle of a long line of games on both sides, so a player would need a minute or more to run around to the other side and catch the other player. And the carnival is extremely loud so that communication between players is impossible.

Once the carneyman has explained the game, he drops a panel down, so all you can see is your selector switch. At the other side, a panel is dropped so that player 2 can only see their selector. You and the other player are instructed to select “up” or “down”. Once you both make your selection, you both drop the marbles into your starting points, and they will roll down to the scoring board, and where they land determines how much money each player wins.

EQUIVALENCY:

This game is a mechanical equivalent to the Prisoner’s Dilemma (PD) scenario in game theory. The only mathematical difference is the payoffs are dollar amounts (positive) rather than years in prison (negative), therefore, in the Prisoner’s dilemma, the player goals are to minimize their prison time, whereas in the interlocked marble race, the goal is to maximize the dollar amount of your prize.

The interlocked marble race creates the same dilemma in the Prisoner’s Dilemma game. The “up” selection chooses the $0/$20 option, and is equivalent to “betray” in PD. The “down” selection chooses the $10/$30 option, and is equivalent to “cooperate” in PD.

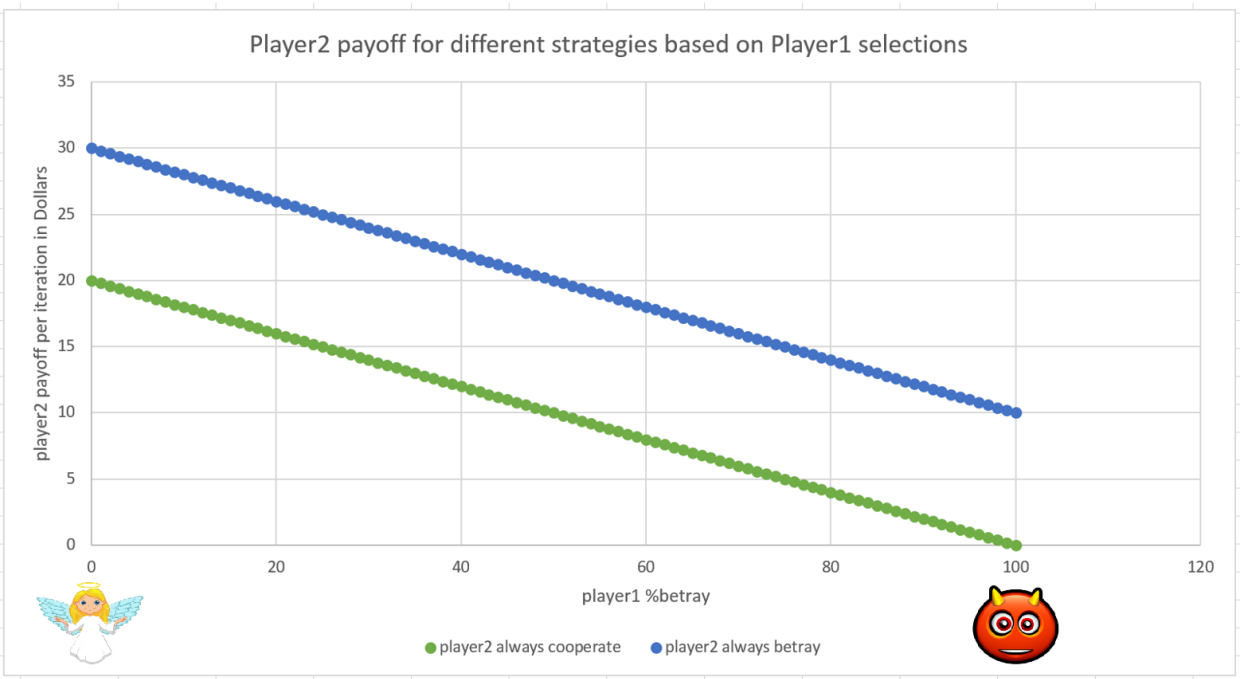
If both Alice and Bob cooperate, both get $20. However, if Alice cooperates, then Bob can get $30 if he betrays, and Alice gets $0. This means that cooperating when the other player betrays you is the worst possible outcome for you, and betraying when the other player cooperates is the best possible outcome for you. So, there is a large incentive to betray, and there is a disincentive to cooperate. However, if both players betray, then they both get the second worst possible outcome, a $10 prize. And thus the dilemma.

STRATEGY:

If our goal is to get ourselves the largest payoff possible, without regard to the other player, what should our strategy be?

We (player2) know nothing about what player1 will select. We could, however, model player1 as a random number generator, with various probabilities of cooperating or betraying, and then we could see what the payoff is for all the different probabilities. Different types of players would have different probabilities for cooperate versus betray. Angels would always cooperate. Devils would always betray. Characters in between have various probabilities. For each character, we could graph the payoff for the other player depending on whether they always cooperate or always betray.

That graph would look something like this:



The X-axis shows player1 (the other player) based on the percent chance they will betray. On the far left, player1 has zero percent chance they will betray us (in other words, they always cooperate, a perfect angel). On the far right, player1 has 100% chance they will betray us (in other words, they always betray, the devil himself).

The Y-axis shows the payoff we (player2) gets for different strategies. The green line shows player2's payoff if we always cooperate. The blue line show's player2's payoff if we always betray. Regardless of what we predict player1 will do, player2's best strategy is to always betray.

For example, let's say Player1 cooperates 20% of the time and betrays 80% of the time, naughty, but not a complete demon. If Player2 always cooperates, they have a 20% chance of winning $20 and an 80% chance of winning $0, which normalizes to $4 per iteration for Player2. If Player2 always betrays, their prize normalizes to $14 per iteration.

Regardless of what kind of character player1 is, angel or devil, we as player2 always gets a higher payoff by betraying rather than cooperating.

ITERATION FALLACY

One of the common misunderstandings of the Prisoner's Dilemma is that our strategy depends on what the other player does. The previous graph shows that regardless of what strategy the other player uses, our best strategy is always to betray.

If we are playing a multiple iteration Prisoner's Dilemma game, then a player can use a tit-for-tat strategy to communicate with the other player based on their previous move: reward them if they cooperate, punish them if they betray. Our choice can be a communication to the other player that we are willing to cooperate if they are.

But this is not an iterated game. There is no previous move to respond to, there is no communication via our current move. The entire game consists of exactly and only one single iteration, and then both players walk away. The strategy that works for an iterated game does not work for a single game.

MORAL FALLACY

Another way players sometimes get confused about the Prisoner's Dilemma is when they look at it as a moral decision instead of an economic decision. An economic decision chooses "what is best for me". A moral decision chooses "what is best for everyone".

If someone looks at the Prisoner's Dilemma or the Interlocked Marble Race with the goal to find the best payoff for everyone, they come up with an answer different from the goal to find the best payoff for themselves. To determine the best payoff for everyone, we have to look at the sum of the prizes both players get.

Player1 Selection	Player2 Selection	Player1 Prize	Player2 Prize	Sum
Cooperate	Cooperate	$20	$20	$40
Cooperate	Betray	$0	$30	$30
Betray	Cooperate	$30	$0	$30
Betray	Betray	$10	$10	$20

When looking at the sum of the prizes, the best outcome comes from both players cooperating. Looking at the Prisoner's Dilemma, or the Interlocked Marble Race, from the moral point of view, from the point of view of what's best for everyone, the solution is for everyone to cooperate, for a maximum sum of $40 for both players.

But the biggest problem with this assumption is that that's not how the game actually works. The players don't sum the prizes and split them in half. The players only get their individual prize and the other player's prize doesn't concern them. Therefore you can choose to cooperate based on the assumption that it is a moral decision to find what is best for everyone, but your assumption doesn't force the other player into viewing the game as a moral decision. They can still look at it from an economic decision, betray your cooperation, and get themselves an even bigger prize.

NARRATIVE DIFFERENCES

The goal of the Interlocked Marble Race was to create a backdrop story that intuitively maps to the payoff matrix given in the game.

I believe one of the difficulties with people understanding the Prisoner's Dilemma is that the narrative of PD creates a backdrop story that intuitively causes some/many people to change the payoff matrix from the one given in the game.

For example, the Prisoner's Dilemma game is a single iteration game, however, the players are criminals in a gang, therefore it isn't unreasonable for someone reading the PD to assume that after being released that the same two prisoner's might be picked up by the police again. That intuitively changes the game to a multiple iterated game, and results in a different solution, because its no longer the single iteration PD.

Another example, the Prisoner's Dilemma explicitly states the payoff matrix for various choices, however, the narrative that creates the backdrop for the story suggests there may be other factors to consider in the actual payoff matrix. In most narratives, both prisoners know each other. In some narratives, they are even members of the same gang. Knowing who the other player is means both players know who betrayed them or cooperated with them, and therefore either player has the capability to seek revenge on the other player if they betray. Gangs sometimes have a "no snitch" policy, which they enforce by punishing any member who rats out other members of the gang. Intuitively, that will modify the payoff matrix from "time in prison" to "time in prison plus or minus whatever punishment someone gets for betraying the other player". At which point, intuitively speaking, it may be wiser to cooperate with the other prisoner, but that's because it's a game different from the Prisoner's Dilemma payoff matrix.

More subtly, the police in the narrative are sometimes presented as being indifferent to justice and only interested in getting a conviction whether the prisoner is actually guilty or innocent. At which point, some will read the narrative and take that their choice should be to oppose the immoral police and that the second prisoner is their ally against these police, and therefore, they cooperate with the other prisoner. That again, is a different game. Having the payoff matrix be enforced by police can make some people intuitively see the police as the agent they are playing against, not the other prisoner.

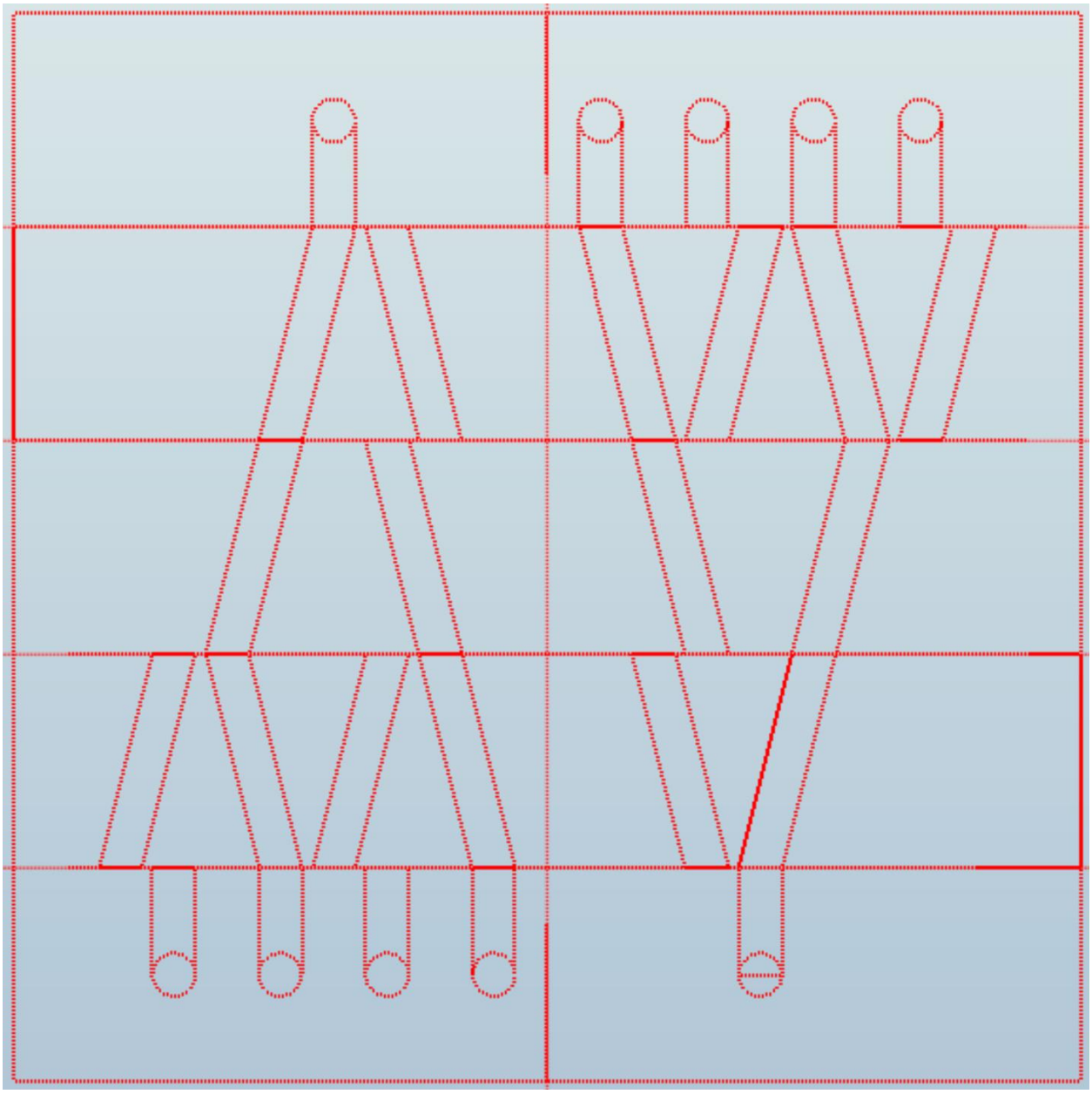
The Interlocked Marble Race tries to use a narrative that intuitively maps to the payoff matrix it describes. It is purely mechanical and the carnival game is not rigged, so there are no evil cops or evil carney people to act against. Players are randomly selected and never know who the other player is, so there is no possibility for repercussions after the single iteration is over. You make your selection, get your prize, and walk away scot free, with the other player never knowing who you are. The players are randomly selected, so getting selected again the next day is unlikely, so a second iteration is unlikely. And even if you did get selected again, the cost of admission is $50, but the max prize is only $30, so you'd lose money overall even if you did get picked again.

Everything about the Interlocked Marble Race narrative was designed to read at an intuitive level to be a single iteration game with no repercussions being imposed after the game, so that it would intuitively read as the explicit game description states.

The goal of the Interlocked Marble Race was to make it easier for readers to intuitively grasp the issue of the Prisoner's Dilemma by having a narrative that intuitively maps to the payoff matrix stated in the game.

HANDS ON: 2D:

Here is a diagram you can print and then cut up to form a puzzle you can move pieces with your hands. Print 2 copies of this page. Take one copy and cut out the second and fourth horizontal bands. They represent the switches. Overlay them on top of the first printout. You can then slide them back and forth and select cooperate or betray.



HANDS ON 3D:

The 2D layout on the previous page is from a 3D printer drawing. The 3D .stl files are available at

http://www.greglondon.com/imr/imr_base.stl

http://www.greglondon.com/imr/imr_slider.stl

This was my first attempt to create a working mechanical version of the puzzle. It has some limitations.

First of all, the base is sized at 4” x 4”. This is small enough to fit on almost all 3D printers (the smallest ones have at least a 4x4x4 print space.) However, this means that the marble tracks are very small, only 5 millimeters wide. A 0.177 caliber BB from a BB gun is 4.5 millimeters in diameter, so it might be just small enough to fit in the marble tracks. This then requires extremely tight tolerances on the puzzle so that the marble doesn’t get stuck in the tracks, so that the marble can cross the gaps between base and slider pieces, and so on. In short, the “marble” (BB) might not work.

The 3D puzzle is still useful, I believe, because it allows a person to hold the mechanism in their hands and see the switches / selectors at work.

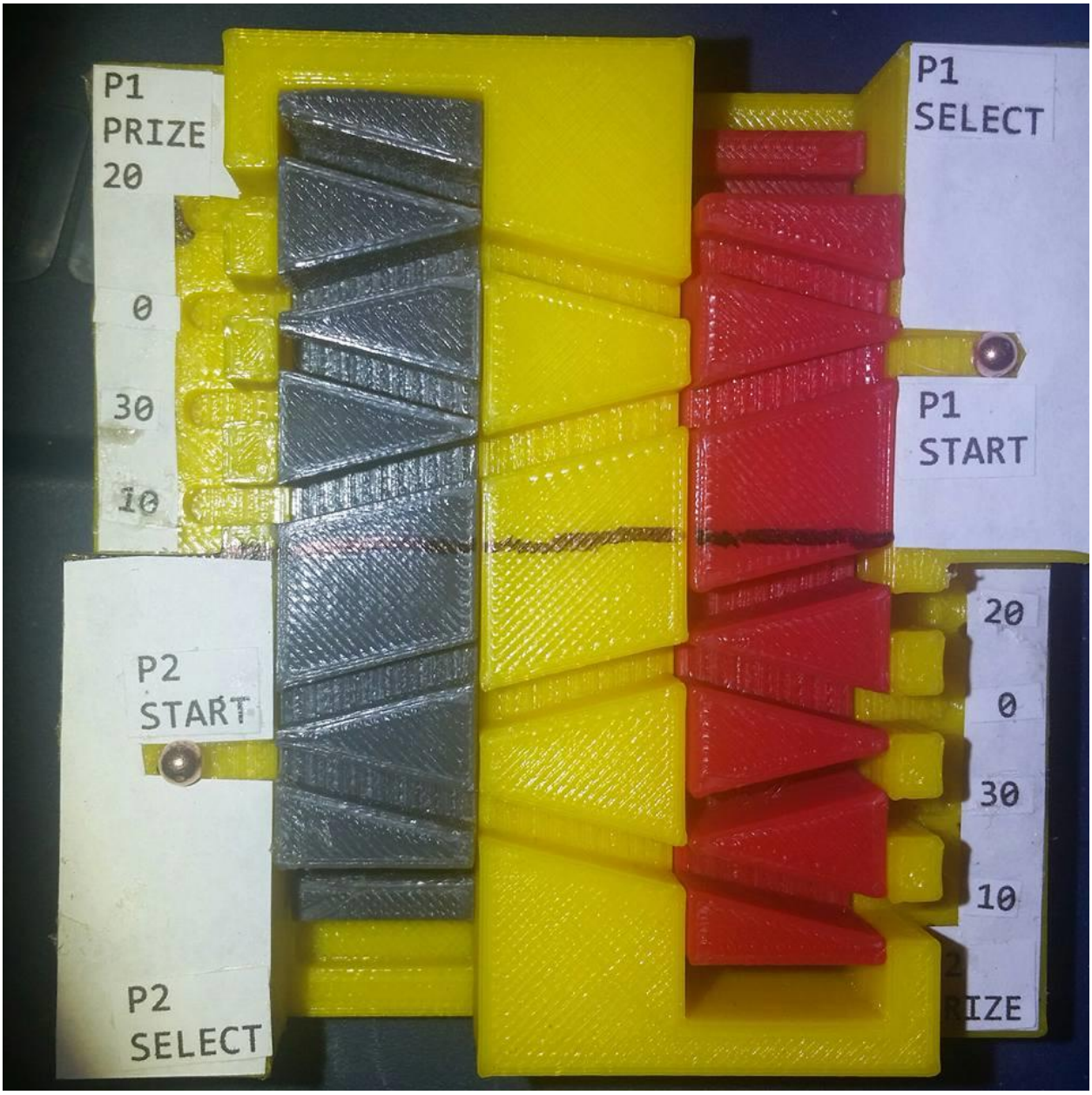
You will need to print 1 base unit and 2 copies of the slider piece. The base requires about 30 meters of filament and about 3.5 hours to print. The slider requires about 8 meters of filament and requires about an hour to print (don’t forget you’ll need to print 2 sliders).

A larger version of the mechanical puzzle is planned. But it’s going to be a bit more complicated because I still want it to be printable on small printers, so it’s going to be a lot more pieces that will have to be fitted together after being printed. I would like it to be large enough to use an actual marble. But that’s a lot of pieces on a 4x4 print bed.

PICTURE OF PUZZLE:

Here is a picture of the 4”x4” 3D printed Interlocked Marble Race. The base is yellow, one slider is red, the other grey. A ruler is included to provide scale. And if you squint, you can see a BB sitting on top of one of the channels. Unfortunately, I haven’t calibrated my 3D printer, so parts are coming out a little small.

Anyway, this picture is just intended to give you a sense of what the parts look like before you print them, so you can better decide if you want to actually use filament.



VIDEO:

I made a video of the Interlocked Marble Race showign the 3D printed puzzle in operation, show how it works, and explain why the best strategy is to betray. You can watch it here:

https://www.youtube.com/watch?v=VV7s5xNG3wQ