

O que é o Dilema dos Prisioneiros

Quando o "equilíbrio" é ineficiente na Teoria dos Jogos

O Dilema dos Prisioneiros é um jogo muito famoso que representa bem o dilema entre cooperar e trair [NOTA 1]. Resumidamente, a estória é a seguinte. Dois suspeitos, A e B, são presos pela polícia. A polícia não tem provas suficientes para os condenar, então separa os prisioneiros em salas diferentes e oferece a ambos o mesmo acordo:



1. Se um dos prisioneiros confessar (trair o outro) e o outro permanecer em silêncio, o que confessou sai livre enquanto o cúmplice silencioso cumpre 10 anos.
2. Se ambos ficarem em silêncio (colaborarem um com o outro), a polícia só pode condená-los a 1 ano cada um.
3. Se ambos confessarem (traírem o comparsa), cada um leva 5 anos de cadeia.

Cada prisioneiro faz a decisão sem saber a escolha do outro - eles não podem conversar. Como o prisioneiro vai reagir? Existe alguma decisão racional a tomar? Qual seria a **sua** decisão?

Usando uma matriz como recurso visual

Uma forma esquemática para mostrar uma interação humana, ou seja, um jogo, é através de uma "**matriz de resultados**" [NOTA 2]. Embora o enunciado do problema seja simples e intuitivo para entender de forma verbal, a representação gráfica oferece uma grande ajuda para visualizar o cenário de forma completa e entender as opções e implicações para cada jogador.

		Prisioneiro A	
		Colaborar (silêncio)	Trair (confessar)
Prisioneiro B	Colaborar (silêncio)	1 ano 1 ano	Livre 10 anos
	Trair (confessar)	10 anos Livre	5 anos 5 anos

Nesta figura você visualiza as duas opções de cada prisioneiro e o resultado de cada combinação de ação. Para cada célula, os valores **vermelhos** a direita referem-se ao **Prisioneiro A**; os **azuis** a esquerda referem-se ao **Prisioneiro B**. Estão descritos quantos anos cada prisioneiro ficará na cadeia. Neste cenário, quando menor o valor da pena, melhor para o prisioneiro.

Os prisioneiros não podem combinar a decisão (estão em salas isoladas e sem comunicação) e devem escolher

simultaneamente. Cada jogador quer ficar preso o menor tempo possível, ou seja, maximizar seu resultado individual. Qual a melhor decisão? Considerando os incentivos deste jogo (os valores das penas para cada combinação de decisões na matriz), existe uma única decisão racional a tomar: **trair**. A explicação é a seguinte:

Imagine que você é o prisioneiro A. Assim, você raciocina nas duas hipóteses:

- Suponha que o Prisioneiro B escolha Colaborar. Então, se você escolher Colaborar, leva 1 ano de prisão. Se escolher Trair, sai livre. Neste caso, Trair é a melhor opção.

- Suponha que o Prisioneiro B escolha Trair. Então, se você escolher Colaborar, leva 10 anos de prisão. Se escolher Trair, fica com 5 anos. Neste caso, Trair é a melhor opção.

Perceba que Trair é a melhor opção em ambos os casos. Em outras palavras, **Trair é a melhor opção independente da decisão do Prisioneiro B.**

Agora, imagine o que o Prisioneiro B está pensando: se ele é racional como você, provavelmente a mesma coisa.

- Ele supõe que você vai escolher Colaborar. Então, se ele escolher Colaborar, leva 1 ano de prisão. Se escolher Trair, sai livre. Neste caso, Trair é a melhor opção.

- Ele supõe que você vai escolher Trair. Então, se ele escolher Colaborar, leva 10 anos de prisão. Se escolher Trair, fica com 5 anos de prisão. Neste caso, Trair é a melhor opção.

De novo, perceba que Trair é a melhor opção em ambos os casos.

O dilema: a escolha individual não é o melhor para ambos

Em Teoria dos Jogos, chamamos que Trair é a **Estratégia Dominante**, ou seja, aquela que apresenta o melhor resultado *independente da decisão do outro jogador*. Quando em um certo jogo, devido o esquema de incentivos (a matriz de resultados) você não precisa se preocupar com a decisão alheia porque existe uma opção melhor independente do seu competidor, então você deve escolher a estratégia dominante.

Neste exemplo dos prisioneiros, como ambos vão escolher Trair devido a estratégia dominante, cada um é preso por 5 anos. Assim, dizemos que Trair-Trair é a **solução de equilíbrio, equilíbrio do jogo** ou **Equilíbrio de Nash**. O Equilíbrio de Nash [NOTA 3] é a solução (combinação de decisões) em que nenhum jogador pode melhorar seu resultado com uma ação unilateral. Ou seja, dado que Trair-Trair é a solução de equilíbrio (o resultado racional do jogo), se o Prisioneiro A mudar unilateralmente para Colaborar ele sai perdendo (15 anos), o mesmo ocorrendo para o Prisioneiro B.

O grande problema no Dilema dos Prisioneiros é que o equilíbrio (Trair-Trair) não é o melhor resultado pois existe um outro possível e melhor: se ambos escolherem Colaborar (ficar em silêncio) cada um ficaria com apenas um ano de prisão. Assim, o Dilema dos Prisioneiros é uma abstração de situações comuns onde a escolha do melhor individual conduz à traição mútua, enquanto que a colaboração proporcionaria melhores resultados. Por isso dizemos que o Dilema dos Prisioneiros resulta em um "**equilíbrio ineficiente**" pois o esquema de incentivos e racionalidade induz a um resultado pior.

Você poderia imaginar que este equilíbrio só ocorre porque as pessoas **não podem conversar e combinar as ações**, e que se pudessem fazer um acordo prévio, tudo se resolveria. Isso não é necessariamente verdade. Você quer colaborar (ficar em silêncio), mas quem garante que o seu parceiro fará o mesmo? O quanto você confia no outro jogador?

Você é o prisioneiro e sua vida está em jogo. Você combina antes que vai colaborar e quer cumprir sua palavra. Seu comparsa sabe isso. Então, o que garante que, no último instante, ele não vai te trair, justamente sabendo que você vai colaborar? Para ele é simples, ele sai livre e você pega 15 anos de prisão... Daí é tarde.

Provavelmente, o seu comparsa pensará da mesma forma a seu respeito. Por isso, o Dilema dos Prisioneiros se torna, na verdade, num **Dilema da Confiança**. Como resolver esse dilema?