

# **ALTRUÍSMO RECÍPROCO**

## **SUA ORIGEM BIOLÓGICA**

Palestra na UFMG e  
Observatório da Piedade

Alaor Chaves – UFMG

## Altruísmo recíproco

Como vimos, pela lógica dos genes, ao altruísmo de parentes faz pleno sentido: ao se sacrificar em benefício de parentes, principalmente em benefício de filhos, o agente (animal ou ser humano) age contra o próprio interesse, mas mesmo assim pode estar favorecendo a transmissão dos seus genes, pois parte deles está presente no genoma dos favorecidos. Por isso, o processo evolucionário darwiniano levou à fixação de genes que geram afeição aos parentes e disposição de se sacrificar por eles. Neste artigo, explicaremos a origem evolucionária do altruísmo entre pessoas não ligadas por parentesco, chamado **altruísmo recíproco**. Como veremos, a cooperação resultante desse é adaptativa no sentido darwiniano, ou seja, ela aumenta o sucesso reprodutivo dos indivíduos cooperativos.



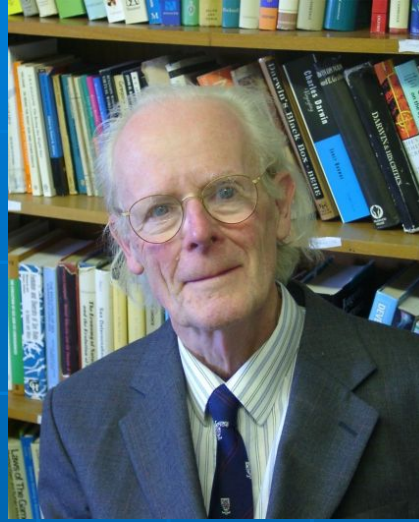
Macacos vervet são cooperativos,  
mas agressivos com os que se comportam  
de forma egoísta

Foto: Stewart Kendall/Allstar

A cooperação (altruísmo recíproco) é comum nos animais sociais. Uma manifestação comum dessa cooperação é a reciprocidade “um te coço, tu me coças”, muito usada para descrever a cooperação entre animais. A compreensão da origem evolucionária do altruísmo entre animais não ligados por laços de sangue (de genes) mostrou ser enormemente desafiadora e requereu a análise de *insights* sobre a evolução por meio da teoria matemática dos jogos.



**John von Neumann**



**John Maynard Smith**



**Robert Trivers**

A teoria dos jogos foi desenvolvida pelo gênio matemático John von Neumann (1903 – 1957) no início dos anos 1940. Ela estuda estratégias usadas por agentes em situações em que seus interesses sejam conflitantes e nas quais o retorno da decisão de cada um depende da decisão de outro(s), que ele não é capaz de prever.

Em 1971, **Robert Trivers** desenvolveu a teoria do altruísmo recíproco. Segundo a teoria, animais que interagem com frequência podem ambos ter ganhos caso cooperem (se sacrificarem um pelo outro). Para que o procedimento funcione, do ponto de vista da evolução biológica, é preciso que os animais também desenvolvam a capacidade de reconhecer os que não retribuem e a partir de então discriminá-los. Para entender a razão disso, consideremos um grupo de animais cuja genética os predisponha ao altruísmo. Cada um concede favores, que têm algum custo e portanto reduzem a capacidade do agente transmitir seus genes.

Se o benefício (medido em termos de aumento do sucesso reprodutivo) recebido for maior do que o sacrifício (medido em termos da diminuição do sucesso reprodutivo), essa transação, considerada isoladamente, aumenta na geração posterior a frequência do gene do altruísmo. Na verdade, em geral há vários genes que predisõem a um dado comportamento, mas por simplicidade consideraremos apenas um gene. Até aqui, tudo bem. Mas considere que apareça no grupo um mutante egoísta, ou seja, desprovido do gene do altruísmo. Ele receberá benefícios dos companheiros, o que aumentará seu sucesso reprodutivo, sem realizar nenhum sacrifício, ou seja, sem nenhuma redução em seu sucesso reprodutivo. Portanto, seu sucesso reprodutivo será superior ao dos colegas altruístas, o que fará com que o gene que o predispõe ao egoísmo se alastre cada vez mais dentro do grupo. Agora já não está tudo bem.

Trivers argumentou que para que o gene do altruísmo se fixe é necessário que os membros do grupo reconheçam os egoístas e adotem estratégias em que eles sejam discriminados e deixem de receber benefícios. Para isso, obviamente, os animais precisam conviver em sociedade, aprender a reconhecer uns aos outros e se lembrar dos companheiros que não retribuem benefícios.

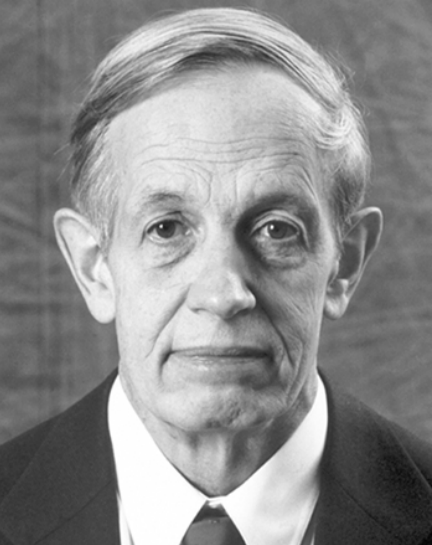
As estratégias, em animais que não desenvolvem cultura, são comportamentos automáticos (instintivos). Instintos operam por meio de sentimentos. Nesse caso, os animais têm de ter genes que os fazem sentir simpatia pelos altruístas e antipatia pelos egoístas.

Em 1973, Maynard Smith introduziu a teoria dos jogos na análise da evolução biológica. Seu intuito era compreender o comportamento social do animal como um conjunto de estratégias para as interações. Sua maior contribuição nesse programa foi introduzir as chamadas **Estratégias Evolutivamente Estáveis (EEE)**. Quando um grupo de animais adota uma EEE, nenhuma outra estratégia adotada por algum dissidente, ou poucos dissidentes, torna seu(s) praticante(s) mais bem sucedido(s) (no sentido darwiniano, ou seja, reprodutivamente mais bem sucedidos) do que os companheiros. Nesse cenário, a estratégia, quando adotada pelo grupo, é vencedora sobre qualquer outra estratégia alternativa adotada por uma pequena minoria. Nesse caso, caso surja no grupo um mutante que adote outra estratégia, seu sucesso reprodutivo será inferior ao dos companheiros e, portanto, o gene mutante não conseguirá se alastrar.



O altruísmo recíproco de Trivers foi investigado em várias espécies animais. Um dos estudos mais importantes, certamente o mais citado, é o altruísmo de morcegos vampiros . Nem todos os morcegos vampiros são bem sucedidos nas caçadas noturnas. Se um morcego fracassa por duas noites consecutivas pode correr risco de vida. Os morcegos desenvolveram um altruísmo recíproco para evitar esse tipo de fatalidade. Ao retornar a casa, morcegos muito bem sucedidos regurgitam um pouco de sangue na boca de algum membro do grupo que tenha fracassado. Essa reciprocidade satisfaz perfeitamente a condição imposta por Trivers. O benefício recebido supera (em muito) o sacrifício feito pelo altruísta.

Veja que invenção magnífica! Sangue é altamente perecível e morcegos não têm refrigeradores. Acabaram encontrando uma maneira de armazenar sangue excedente: na cabeça de algum membro do grupo.



**John Forbes Nash**

Embora a teoria do altruísmo recíproco tenha sido desde o início amplamente aceita, para que ela se tornasse mais bem fundamentada na teoria dos jogos muito trabalho técnico teve de ser realizado. Smith demonstrou a existência das EEE. Na verdade a EEE é uma versão, para o caso da evolução biológica, do chamado **Equilíbrio de Nash**, descoberto pelo matemático **John Nash (1928 – )** em 1950. Nash aplicou a teoria dos jogos a interações entre agentes econômicos.

Cada agente tem conhecimento das estratégias dos demais e é tentado a usar estratégias que maximizem os seus ganhos. No conjunto as estratégias dos agentes podem gerar um estado de equilíbrio. Nesse estado, nenhum agente é capaz de aumentar os seus ganhos mudando solitariamente a sua estratégia. Se todos os agentes, ou boa parte deles, adotassem uma estratégia de colaboração, todos poderiam sair ganhando. Mas sem esse tipo de pacto em que as estratégias sejam alteradas de maneira conjunta, todos os agentes ficam aprisionados em um estado em que, relativamente, todos saem perdendo. Nash acabou ganhando o Nobel de Economia em 1994 pelo trabalho. No caso do Equilíbrio de Nash, os agentes podem pactuar outras estratégias. Essa solução teleológica é impossível na evolução biológica, na qual o comportamento é condicionado pelos genes.



John Nash e Robert Trivers são dois gênios vitimados pelo mesmo infortúnio, a esquizofrenia. Nash ingressou como estudante em Princeton com uma carta de recomendação que dizia secamente: “Esse homem é um gênio.” Logo ao iniciar os estudos, os sintomas começaram a aparecer. Nos momentos de lucidez, criou várias coisas, dentre elas a teoria que lhe rendeu o Nobel e que os matemáticos não incluem dentre os seus maiores trabalhos. Sua vida foi narrada no filme **Uma Mente Brilhante**.

Trivers ingressou em Harvard para estudar matemática, mas se graduou em história. Não foi aceito em Yale para estudar Direito por causa da sua doença. Vagou de uma atividade para outra, até que retornou a Harvard, onde em 1972 se doutorou em biologia. Nos anos 1971 – 1976, deu contribuições seminais para a teoria da evolução. É considerado o maior teórico vivo em evolução. Steven Pinker o considera um dos maiores pensadores da história do ocidente.

# Dilema do Prisioneiro

O dilema do prisioneiro é talvez o mais investigado dos jogos. Seu interesse reside no fato que ele é simples e exibe com toda a clareza a dificuldade de se chegar à colaboração. Mas seu estudo detalhado mostrou que a colaboração pode resultar do jogo cego da evolução biológica.

Considere dois prisioneiros acusados de praticar um dado crime. As evidências só provam delitos para os quais a pena é de 1 ano, mas o juiz sabe que há algo mais grave e quer condenar alguém a pena mais longa. Conversa com cada um deles em separado e lhe promete redução da pena para 6 meses caso ele traia o parceiro testemunhando contra ele, levando-o à pena de 4 anos. Se eles se traírem mutuamente, cada um levará pena de 2 anos. O dilema é atroz. É tentador trair e cada um sabe que o parceiro será também tentado. Nesse caso, se ele colaborar com o parceiro (for leal) e for traído, levará 4 anos de prisão. A lógica pura levará ambos à traição e, portanto, a penas individuais maiores do que a obtida com a colaboração, que é de 1 ano. Os números são arbitrários e podem ser alterados na medida da perversidade do enunciador.

Para análise matemática, é conveniente reformular o dilema em termos de ganhos entre os dois jogadores. Sejam os seguintes os ganhos do jogo:

Se A trai e não é traído por B, A ganha 5 pontos e B ganha 0 pontos;

Se A não trai nem é traído por B, ambos ganham 3 pontos;

Se ambos traem, ambos ganham 1 ponto.

Supõe-se que ambos os oponentes só visarão o interesse pessoal. Nesse caso, a melhor estratégia é trair e, como ambos a seguirão, ambos ganharão 1 ponto. Se eles pudessem combinar as estratégias e confiar um no outro, ambos ganhariam 3 pontos. Um jogador leal que jogar uma única vez com um traidor sairá sem nenhum ponto do jogo. Mas o jogo pode ser elaborado em dois aspectos:

1)O jogo pode ser repetido e cada jogador saberá o que o adversário fez nas jogadas passadas.

2)O jogo não só pode ser repetido, como também cada jogador enfrentará diferentes adversários com estratégias possivelmente distintas.

Estamos agora falando no **dilema do prisioneiro iterativo** (DPI)

Suponhamos inicialmente o jogo iterativo entre um egoísta (E), cuja estratégia é sempre trair, e um altruísta (A), que nunca trai. Em cada iteração do jogo, E ganhará 5 pontos e A ganhará 0 pontos. Se jogarem  $N$  vezes, E acumulará  $5N$  pontos, contra 0 pontos acumulados por A. É um verdadeiro massacre. Altruístas não têm qualquer chance jogando apenas contra egoístas. Fato semelhante ocorrerá se um grupo de  $N$  de altruístas jogar contra  $N$  egoístas. Para simplificar, suporemos que não haja iteração, ou seja, cada jogador jogará uma única vez contra todos os outros. Vejamos o resultado.

Cada A jogará contra os  $N$  E e somará 0 pontos.  
Jogará também contra os  $N - 1$  A, no que somará  $3(N-1)$  pontos.  
O total de pontos de cada A será então  $3(N-1)$ .

Cada E jogará contra os  $N$  A e somará  $5N$  pontos.  
Jogará também contra os  $N - 1$  E, no que somará  $N - 1$  pontos.  
O total de pontos de cada E será portanto  $6N - 1$ .

Novamente, haverá um massacre. Haverá alguma estratégia na qual  $N$  jogadores possa levar vantagem sobre  $N$  egoístas?

Em 1980, **Robert Axelrod**, da Rand Corporation, realizou um torneio com o propósito de investigar esta questão. O torneio seria entre programas de computador, cada um com sua estratégia. Programadores foram convidados a elaborar programas e participar do torneio. O vencedor foi o programa **Tit for Tat** (TFT), formulado por **Anatol Rapoport**. Tit for Tat é uma expressão inglesa para retaliação equivalente. É o olho por olho, dente por dente, do Velho Testamento. A estratégia do TFT é surpreendentemente simples. Em um jogo iterativo contra um dado jogador, na primeira rodada ele é sempre altruísta, ou seja, não trai. Na sequência, em cada rodada ele faz o que o adversário fez na rodada anterior. Para exemplificar, consideremos um TFT jogando  $N$  rodadas contra um E. Na primeira rodada, TFT ganha 0 pontos e E ganha 5. Em cada uma de todas as  $N - 1$  rodadas posteriores, cada um deles ganha 1 ponto. Assim:

TFT soma  $N - 1$  pontos,

E soma  $5 + N - 1$  pontos, ou seja,  $N + 4$  pontos.

Isso mostra que 1 TFT não é capaz de vencer individualmente 1 egoísta. Mas quando há no torneio outros tipos de jogadores, TFT sai vencedor. A vitória de TFT no torneio de Axelrod surpreendeu a todos. O torneio foi repetido e programas mais elaborados participaram, com estratégias muito inteligentes para enganar os adversários.

Novamente, TFT foi o campeão.

Nos torneios, cada jogador enfrentava cada um dos adversários em partida de 200 rodadas e o campeão seria o que somasse mais pontos durante o torneio. Desde então, vários torneios foram realizados e TFT nunca foi vencido. Acredita-se que ele seja invencível e há matemáticos que buscam a demonstração matemática dessa invencibilidade. Axelrod faz uma análise do dilema do prisioneiro e o que ele pode nos ensinar sobre a evolução do comportamento animal no livro *The Evolution of Cooperation* (Edição revisada 2006).

Podemos perceber um pouco da eficácia da estratégia TFT considerando um tipo mais simples de torneio. Nele competem 2 TFT (TFT1 e TFT2) e 2 egoístas (E1 e E2). Cada participante joga N vezes contra todos os adversários. Naturalmente, TFT1 somará o mesmo número de pontos de TFT2 e E1 somará o mesmo número de pontos de E2.



TFT1:

N jogos com E1. Soma parcial de pontos =  $N - 1$

N jogos com E2. Soma parcial de pontos =  $N - 1$

N jogos com TFT2. Soma parcial de pontos =  $3N$

$S(\text{TFT1}) = 5N - 2$ .

E1:

N jogos com TFT1. Soma parcial de pontos =  $5 + N - 1$

N jogos com TFT2. Soma parcial de pontos =  $5 + N - 1$

N jogos com E2. Soma parcial de pontos =  $N$

$S(\text{E1}) = 3N + 8$ .

A diferença entre os dois desempenhos será

$$D = S(\text{TFT1}) - S(\text{E1}) = 5N - 2 - (3N + 8) = 2N - 10.$$

A estratégia TFT é vencedora se  $N$  for maior do que 5.

Suponhamos que mais 18 egoístas sejam adicionados ao jogo. Então, teremos 22 jogadores, 2 TFT e 20 E. Fazendo as contas dos resultados dos jogos, teremos:

$$S(\text{TFT1}) = 23N - 20$$

$$S(\text{E1}) = 21N + 8.$$

A diferença entre os dois desempenhos será

$$D = 23N - 20 - (21N + 8) = 2N - 28.$$

Agora, a estratégia TFT será perdedora se N for menor do que 14 e vencedora se N for maior do que 14. Quanto maior o número de E participantes do jogo, mantidos somente 2 TFT, maior tem de ser o número N de iterações para que a estratégia TFT vença a estratégia egoísta. Mas no longo prazo ela sempre vence.

O que isso pode nos ensinar sobre a evolução do altruísmo?

Como argumentou Trivers, para que o altruísmo recíproco possa evoluir em um grupo de animais sociais, é necessário que os indivíduos altruístas reconheçam os que não colaboram em reciprocidade e cessem de lhes prestar benefícios. A estratégia TFT inclui esse princípio. Vimos que em um grupo de animais sociais (que interagem com frequência) praticantes de TFT conseguem maior soma de benefícios do que os egoístas. Se não for solitário no grupo, o indivíduo TFT pode levar vantagem sobre qualquer egoísta, mesmo que os egoístas sejam muito mais numerosos, desde que as interações entre eles sejam suficientemente abundantes. Essa vantagem resultará em maior sucesso procriativo dos TFT e portanto em aumento dessa variedade no grupo.

A estratégia TFT contém um elemento importante não considerado por Trivers. O indivíduo IFT retalia a falta de colaboração na mesma intensidade, mas tão logo o outro agente colabora ele retoma a colaboração. Essa estratégia vence outras mais retaliadoras, por exemplo duas traições em retaliação a uma traição. Esse é um fato de origem matemática com consequências muito positivas para as sociedades animais. Os “rancorosos”, que se negam a colaborar com quem foi egoísta mas já não é, não são, do ponto de vista darwiniano, os mais bem adaptados.

O altruísta incondicional também não é adaptado. Sua existência é um fato, mas sua origem é provavelmente cultural.

## O que isso nos pode dizer sobre o comportamento humano?

Não se pode entender o comportamento humano considerando-se somente os genes. Uma das características mais importantes do homem é sua tendência e capacidade para criar **cultura**. Naturalmente, essa característica tem base nos genes. Ou seja, a natureza humana contém instintos que influenciam o seu comportamento, mas também contém instintos e capacidades que levam à criação de cultura, e cultura também influencia o comportamento. O comportamento humano é resultante tanto da sua natureza quanto da cultura em que ele se desenvolveu. Criação de cultura não é uma capacidade exclusiva do Homem. Muitos outros animais sociais também desenvolvem culturas. Mas no Homem a cultura adquiriu uma dimensão incomparavelmente maior. A questão do peso relativo dos dois fatores, o genético e o cultural, no comportamento humano tem sido objeto de antiga e interminável controvérsia. Nos tempos modernos, **John Locke** introduziu a noção da **tabula rasa**, segundo a qual a mente de um recém-nascido é uma página em branco e tudo o que se desenvolve nela é escrito pela experiência.

Como antítese à *tabula rasa*, no final do século 19 e início do século 20 alguns outros pensadores passaram a defender um determinismo genético pelo qual as propensões e aptidões humanas são genéticas e imutáveis. Criou-se assim a falsa dicotomia *nature versus nurture* (natureza versus criação) introduzida por Francis Galton. Essa polarização dominou a história do século 20, que foi o Século da Ideologia. Investigações científicas menos apaixonadas levaram a quase totalidade dos biólogos, psicólogos evolucionários, geneticistas do comportamento, neurocientistas e até mesmo linguistas a entenderem melhor a natureza humana e o comportamento humano como expressões dos genes em dada cultura. Mas muitos humanistas, antropólogos, cientistas sociais e educadores ainda não estão convencidos e continuam defendendo a ideia da *tabula rasa*. Para eles, não há nenhum sentimento, nem mesmo o amor paterno, nem nenhuma aptidão (por exemplo para matemática ou música ou jogar futebol) que tenha origem genética. Defendem o **determinismo cultural**. Nosso ponto de vista é o dos cientistas naturais.



Nos jogos mencionados anteriormente, as estratégias são instintos que os jogadores seguem deterministicamente. Elas podem incorporar aprendizado em um sentido limitado: a estratégia pode ser um algoritmo elaborado que leva em conta todas as jogadas passadas para decidir sobre o que fazer na próxima. Mas há uma variedade de estratégias, ainda não considerada, que leva em conta a experiência prévia de outra maneira. O programa pode conter várias estratégias que se vêm e uma dada estratégia pode se deletar em benefício de outras. Pelo menos um programa desse tipo venceu o TFT. Essa elaboração torna o jogo mais próximo da realidade humana, que contém genes e também cultura. Nos jogos das interações sociais, os seres humanos podem também fazer pactos entre parte do grupo ou em todo ele. Por exemplo, nos jogos econômicos parte dos agentes podem formar cartéis.

Entretanto, os ancestrais humanos tiveram de colaborar antes de desenvolver culturas mais complexas. Se retrocedermos o bastante, chegaremos a ancestrais em que a cooperação era baseada apenas em instintos. Os instintos que geraram colaboração certamente ainda influenciam grandemente o nosso comportamento. Além do mais, o que veio depois, as sociedades complexas que criamos, são consequência desse nosso processo evolutivo.

Boa parte da moralidade humana tem origem nos instintos e capacidades associados ao altruísmo recíproco. Podemos citar:

Nossa simpatia pelas pessoas altruístas e antipatia pelos egoístas

Nosso sentimento de gratidão

Nosso sentimento de equidade (distribuição adequada dos benefícios )

Recompensas por ações muito benéficas à sociedade e sanções a ações muito maléficas

Nosso sentimento de justiça e indignação frente à sua violação

Nossa habilidade para reconhecer trapaceiros.

Esses atributos são **Universais Humanos** (presentes em todas as culturas). Portanto, não são culturais.

Tem sido amplamente apontado o fato de que, ao realizar um ato altruísta, as pessoas acham importante que o grupo ao qual pertencem fique sabendo. Quem faz uma grande caridade em geral quer que ela se torne pública. Atribui-se isso à vaidade, mas este sentimento é derivação de outros sentimentos adaptativos. É decorrência do instinto de altruísmo recíproco que as pessoas considerem importante o grupo saber que elas deram uma contribuição importante. Se o indivíduo faz um feito heroico, todos devem saber! Se realiza um grande feito intelectual (o que contribui para o bem do grupo), todos devem saber! A reputação indivíduo é na verdade a soma dos pontos que ele somou colaborando para o grupo. Se as pessoas não zelassem pela reputação, a sociedade humana não teria chegado ao que é.

No altruísmo recíproco ambos os agentes acabam ganhando, ou seja, obtendo benefícios. Portanto, ele é meramente uma estratégia vencedora e não um altruísmo no sentido rigoroso do termo. O altruísmo recíproco é uma troca em que as partes só visam o seu interesse, nada mais do que isso. Mas essa estratégia é vencedora em confronto com qualquer outra.

Gratos, matemática! Sem sua lógica de ferro, sem a qual não teríamos evoluído, não estaríamos aqui. Muito menos seríamos propensos à colaboração. Mais ainda: **a base da moralidade humana é fruto do jogo cego da evolução biológica, é gerada pela leis da probabilidade.**