IERFH
Instituto Ética, Racionalidade
e Futuro da Humanidade

...O que é um viés, mesmo?

por **Eliezer Yudkowsky**, 2006*

Um *viés* [*bias*] é um certo tipo de obstáculo para o nosso objetivo de alcançar a verdade – sua característica de ser um obstáculo vem por termos por objetivo a verdade – mas há muitos obstáculos que não são "vieses".

Se nós começarmos perguntando "O que é um viés?", ele surge na questão na ordem errada. Como diz o provérbio, "Existem quarenta tipos de loucura mas apenas um tipo de senso comum". A verdade é um alvo estreito, uma pequena região a ser acertada do espaço de configuração. "Ela me ama, ela não me ama" pode ser uma questão binária, mas $E=mc^2$ é um pequeno ponto no espaço de todas as equações, como um bilhete vencedor de loteria no espaço de todos os bilhetes de loteria. O erro não é uma condição excepcional; é o sucesso que é *a priori* tão improvável que requer uma explicação.

Nós não começamos com um dever moral de "reduzir os vieses", porque vieses são maus e Não São Permitidos. Essa é a maneira de pensar à qual alguém pode ser levado se adquirir um dever deontológico de 'racionalidade' por osmose social, o que leva as pessoas a tentarem executar técnicas sem apreciar a razão de ser delas. (O que é ruim e mal e Não É Permitido, de acordo com *O sr. está brincando, Sr. Feynman?*, que eu li quando era criança.)

Em vez disso, nós começamos querendo chegar à verdade, por seja lá qual for a razão, e nós encontramos vários obstáculos que se colocam no caminho de nosso objetivo. Esses obstáculos não são completamente dissimilares uns dos outros – por exemplo, há obstáculos que têm a ver com não ter poder de computação o suficiente disponível, ou com as informações serem caras. Acontece, de fato, que um grande grupo de obstáculos parece ter uma certa característica em comum – se

aglomerando em uma região do espaço de 'obstáculos para a verdade' – e esse aglomerado foi rotulado de "vieses".

O que é um viés? Nós podemos olhar para o aglomerado empírico e encontrar um teste compacto para decidi-lo? Talvez nós descubramos que nós não podemos dar nenhuma explicação melhor do que apontar para alguns exemplos excepcionais, e torcer para que o interlocutor entenda. Se você é um cientista que acabou de começar a investigar o fogo, pode ser muito mais sábio apontar para uma fogueira e dizer "Fogo é essa coisa laranja, brilhante e quente que está ali", em vez de dizer "Eu defino fogo como uma transmutação alquímica de substâncias que libera flogisto". Como eu disse em **A Simples Verdade**, você não deveria ignorar algo só porque você não é capaz de defini-lo. Eu não posso citar de cabeça as equações da Relatividade Geral, mas de qualquer modo eu sei que, se eu andar para além de um penhasco, eu vou cair. E nós podemos dizer o mesmo de vieses – eles não vão nos atingir nem um pouco menos duramente se ocorrer que nós não podemos definir o que é um "viés". Então nós poderíamos apontar para uma conjunção de falácias, para confiança excessiva, para a disponibilidade de heurísticas de representatividade, ou para a negligência da taxa básica e dizer, "coisas assim".

Com tudo isso dito, nós aparentemente rotulamos como "vieses" aqueles obstáculos para a verdade os quais são produzidos, não pelo custo de informações, não por poder limitado de computação, mas pela forma de nossa própria maquinaria mental. Por exemplo, a maquinaria está evolutivamente otimizada para propósitos que se opõem ativamente à precisão epistêmica; por exemplo, a maquinaria utilizada para ganhar discussões em contextos políticos adaptativos. Ou a pressão da seleção distorceu a precisão epistêmica; por exemplo, acreditar no que os outros acreditam, para se dar bem socialmente. Ou, no clássico heurística e viés, a maquinaria opera com um algoritmo identificável que faz algum trabalho útil, mas também produz erros sistemáticos: a heurística da disponibilidade não é em si mesma um viés, mas ela dá origem a alguns vieses identificáveis e compactamente descritíveis. Nossos cérebros estão fazendo algo de errado, e depois de muita experimentação e/ou muita reflexão, alguém identifica o problema de uma maneira que o Sistema 2 [isto é, outra pessoa] pode compreender; então nós chamamos isto de "viés". Mesmo que nós não possamos melhorar em nada só por conhecê-lo, ainda assim é uma falha que surge, de uma maneira identificável, de um tipo particular de maquinaria cognitiva – não de ter maquinaria insuficiente, mas sim da própria forma da maquinaria.

"Vieses" são distinguidos de erros que surgem de conteúdo cognitivo, tal como crenças adotadas, ou deveres morais adotados. Nós chamamos esses de "enganos" em vez de "vieses", e eles são muito mais fáceis de corrigir, quando nós mesmos os tivermos percebido. (Apesar de a origem do engano, ou a origem da origem do engano, poder ser, em última instância, um viés.)

Platão não era "enviesado" porque ele não conhecia a Relatividade Geral – ele não tinha nenhuma maneira de coletar essa informação, sua ignorância não surgiu da forma de sua maquinaria mental. Mas se Platão acreditava que filósofos dariam reis melhores porque ele próprio era um filósofo – e se essa crença, por sua vez, surgisse de um instinto político adaptativo universal de autopromoção, e não porque o pai de Platão disse a ele que todos têm um dever moral de promover sua própria profissão para governar; ou ainda porque Platão cheirou muita cola quando criança – então isso seria um viés, ainda que Platão nunca tenha sido alertado disso.

Vieses podem não ser baratos para corrigir. Alguns podem nem mesmo ser corrigíveis. Mas onde nós podemos olhar para nossa maquinaria mental e ver uma explicação causal de uma classe identificável de erros; e quando o problema parece vir da forma evoluída da maquinaria, em vez de vir da maquinaria ser insuficiente, ou de conteúdo específico ruim; então nós chamamos isso de viés.

Pessoalmente, eu vejo nossa missão em termos de aquisição de habilidades pessoais de racionalidade, de aprimoramento das técnicas de descoberta da verdade. O desafio é obter o objetivo positivo da verdade, não evitar o alvo negativo da falha. O espaço de falha é amplo, erros infinitos de infinita variedade. É difícil descrever um espaço tão gigantesco: "O que é verdade para uma maçã pode não ser para outra maçã; por isso mais pode ser dito de uma única maçã do que de todas as maçãs do mundo". O espaço de sucesso é mais restrito e, por conseguinte, mais pode ser dito sobre ele.

Enquanto eu não repudio (como se pode ver) discutir definições, nós devemos lembrar que esse não é nosso objetivo primário. Nós estamos aqui para seguir na grande busca humana pela verdade; pois nós temos uma necessidade desesperada de conhecimento, e, além disso, nós somos curiosos. Para alcançarmos esse fim, nos empenhemos para superar sejam quais forem os obstáculos que estejam em nosso caminho, quer nós os chamemos de "vieses", quer não.

Notas

* Texto traduzido por Lucas Machado. Revisado por Lauro Edison. O original pode ser lido aqui: **http://lesswrong.com/lw/gp/whats_a_bias_again/**