

BEST-SELLER DO  
NEW YORK TIMES

# RUÍDO

Uma falha no julgamento humano

DANIEL  
KAHNEMAN

AUTOR DE RÁPIDO E DEVAGAR

OLIVIER  
SIBONY

CASS R.  
SUNSTEIN

OBJETIVA

aos olhos de  
**TIO HULI E RAIANA DIAS**

# SÚMARIO

Introdução .....	3
Um Sistema Ruidoso .....	9
Decisões Singulares .....	13
Questões de Julgamento .....	15
Medindo o erro .....	18
Análise do Ruído .....	21
Ruído De Ocasão .....	25
Como o grupo amplifica o ruído .....	28
Julgamentos e Modelos .....	32
Regras Sem Ruído .....	35
Ignorância Objetiva .....	37
O Vale do Normal .....	40
Heurísticas, vieses e ruído .....	45
A operação de equiparação .....	50
Escalas .....	54
Padrões .....	58
As fontes do ruído .....	60
Juízes melhores para julgamentos melhores .....	63
Desenviesamento e higiene da decisão .....	66
Sequenciando informações na ciência forense .....	68
Seleção e Agregação em previsões .....	71
Diretrizes na medicina .....	74
Definindo a escala em análises de desempenho .....	78
Estrutura em contratações .....	82
Protocolo de avaliações mediadoras .....	85
Os Custos da Redução de ruído .....	88
Dignidade .....	91

Regras ou padrões .....	93
Revisão e Conclusão .....	97

# INTRODUÇÃO

## *Dois Tipos de Erro*

“Esse livro é sobre erro humano. Viés e ruído - desvio sistemático e dispersão aleatória - são componentes diferentes do erro. Os alvos ilustram a diferença.”

Uma propriedade geral do ruído é que podemos reconhecê-lo e medi-lo mesmo sem saber nada sobre o alvo ou viés. Inúmeras conclusões a que chegamos são extraídas de julgamentos cuja resposta real é ignorada ou incognoscível.

Para compreender o erro no julgamento, devemos compreender tanto o viés como o ruído. O ruído é o problema mais importante, muitas vezes em discussões públicas sobre o erro humano e em organizações pelo mundo, ele nem é reconhecido. O viés rouba a cena. Esse livro traz a tentativa de reestabelecer o equilíbrio.

No mundo real, a quantidade de ruído é muito alta como nos exemplos abaixo. Esses são casos em que a precisão importa:

- Na medicina: vários médicos oferecem diversos diagnósticos diferentes para a mesma doença. Como na psiquiatria em que o julgamento é subjetivo.
- Nas decisões sobre guarda de menores: assistentes sociais de pouco tato tendem a mandar mais crianças

pra adoção em comparação a outros.

- Previsões especializadas como as vendas de um novo produto ou taxa de desemprego.
- Decisão sobre pedidos de asilos: sua aprovação funciona como uma loteria.
- Decisões pessoais dentro de uma empresa: pessoas que realizam entrevistas de emprego avaliam de maneira muito diferente os mesmos candidatos.
- Decisões sobre a fiança de um réu: dependem dos juízes.
- A ciência forense é ruidosa: principalmente em análises de DNA.
- Concessões de patentes: depende do encarregado do pedido.

Sempre que observarmos um julgamento humano, provavelmente encontraremos ruído. Para melhorar o julgamento precisamos eliminar o viés e o ruído.

## **Parte 1- Encontrando o Ruído**

- Crime e ruidoso castigo

Quando alguém é condenado por um crime sua sentença provável não deveria depender do juiz

específico designado para o cargo. Porém no mundo todo, juízes desfrutam de poder arbitrário sobre as sentenças que se baseiam em diversos fatores como o caráter e as circunstâncias do réu e não só o crime.

Nos anos 1970, devido à grande incidência de ruídos, a discricionariedade (arbitrariedade) com que as sentenças eram definidas, começou a desmoronar. Marvin Frankel, um dos fundadores da organização Human Rights First, apresentou uma série de casos convincentes de resoluções de processos criminais definidas a partir das opiniões, inclinações e preferências do juiz.

Frankel afirmou que os juízes federais possuíam poderes quase completamente indiscriminados que resultavam em crueldades por ele consideradas arbitrárias em um governo de leis, não de homens. Ele exortou o congresso a pôr um fim nos ruídos, nas formas de variações inexplicáveis nas sentenças mas também estava preocupado com o viés, demonstrado nas disparidades raciais e socioeconômicas.

Para combater os ruídos e vieses, defendia que deveriam ser realizados testes relevantes e objetivos como perfis detalhados, checklist de fatores, incluindo sempre que possível graduações numéricas. Um processo mais mecânico contendo regras impessoais, aplicáveis a todos, obrigatórias. Também recomendou uma comissão para tratar do assunto. Seu livro se tornou um dos mais influentes da história do direito criminal.

Diversos estudos foram feitos. Um deles dirigido por Frankel avaliou sentenças de réus hipotéticos conferidas por 50 juízes. Como conclusão observou que não havia consenso entre elas e as variações entre as penas eram espantosas. Um traficante de heroína por exemplo, poderia pegar de 1 a 10 anos, roubos a bancos de 5 a 18 anos de prisão, dependendo do juiz.

Por mais reveladores que sejam esses estudos rigorosamente controlados, eles subestimam a magnitude do ruído na justiça criminal. Por exemplo, pesquisas revelam que juízes tendem a dar sentenças mais leves no começo do dia, após uma refeição, em dias de sol... enquanto tendem a pesar mais na sentença depois que seu time do coração perde uma partida.

## **Reduzindo o Ruído nas Sentenças**

Os argumentos de Frankel e resultados desses estudos chegaram ao irmão de John Kennedy, Edward Kennedy, um dos membros mais influentes do senado americano. Em 1975 introduziu uma legislação para a reforma judiciária que só foi para frente em 1984 com a aprovação do Congresso para a Lei da Reforma das Sentenças.

A nova legislação se destinava a acabar com o ruído, restringindo o poder arbitrário dos juízes por meio de diretrizes criadas pela Comissão de Sentenças dos EUA. Antes de determinar a sentença os juízes deveriam considerar o crime e a ficha criminal. Combinando esses

dois critérios, a decisão só poderia variar no extremo superior 25% e no inferior entre 6 meses e 25%. Embora obrigatória, as restrições não eram estritamente rígidas, ainda havia espaço para manobras dos juízes. Estudos demonstram que ainda há ruído porém as diretrizes reduziram a disparidade na variação das sentenças.

Apesar dos resultados as diretrizes foram alvo de críticas, muitos juízes achavam que haviam sentenças severas demais, o que tem a ver com o viés e não o ruído. As diretrizes também não permitiam que os pormenores do caso fossem considerados. *“Faz-se necessidade não de cegueira, mas de uma visão penetrante, de equidade, que só pode ocorrer em um julgamento que leva em consideração as complexidades do caso individual.”*

Em 2005 a Suprema Corte anulou as diretrizes, passou a apresentá-las apenas com caráter orientador. Pesquisas mostraram que essa mudança produziu um aumento significativo de disparidade entre as decisões dos juízes. A obrigatoriedade reduzia tanto o viés quanto o ruído. A diferença entre decisões de sentenças de réus afro americanos e brancos para o mesmo crime também aumentou muito.

Julgar é difícil porque o mundo é incerto, sua complexidade se mostra clara na justiça e em muitas outras áreas profissionais. A discordância é inevitável onde quer que haja julgamento envolvido. A extensão dessa discordância é muito maior que o esperado.

O ruído do sistema gera injustiça generalizada, altos custos econômicos. O ruído pode ser reduzido, regras e diretrizes entre outras estratégias podem ser usadas, que conseqüentemente reduzem o viés. Porém as estratégias de redução podem gerar objeções e muita dificuldade.

## UM SISTEMA RUIDOSO

Em uma grande empresa, muitos profissionais estão em posição de fazer julgamentos que a comprometem. No caso de uma companhia de seguros, por exemplo, corretores fazem cotações de prêmios por riscos financeiros, analistas preveem o custo de futuros resgates... Mesmo com corretores conceituados na equipe, quando há uma solicitação, ela é atendida pelo funcionário disponível, ou seja, o corretor incumbido da cotação é selecionado por uma loteria.

O valor do prêmio da cotação também impacta significativamente a companhia, um prêmio elevado é vantajoso porém mais difícil de ser aceito, correndo o risco de perder o contrato para a concorrência. Um prêmio baixo tem mais probabilidade de ser aceito, porém não traz tanto retorno para a companhia. Existe um preço justo para cada cotação e há uma grande chance de um grupo de profissionais se aproximar dele. Valores muito acima, ou muito abaixo, saem caro para a empresa. É essa variação que prejudica a organização.

A estimativa inicial do analista ao fechar um contrato, irá determinar suas futuras negociações com o segurado. Ofertas baixas geram maiores probabilidades do segurado recorrer a processos na justiça, ofertas altas podem aumentar o poder de barganha do segurado facilitando o analista a concordar com demandas

infundada trazidas por ele já que a companhia é obrigada por lei a reservar o valor previsto em cada solicitação.

A escolha do corretor fica por conta do acaso. A loteria que determina um corretor para determinado caso produz variabilidade, mas ela permanece ignorada. Nas auditorias de ruído, essa variabilidade vem a tona. Para as auditorias, os executivos da companhia produziram 5 casos representativos para cada grupo de corretores e analistas, onde cada funcionário deveria analisar de 2 a 3 casos. Os resultados foram chocantes, a variação média encontrada entre as diferentes cotações foi de 55 %. O custo anual desse ruído para a companhia foi de centenas de milhões de dólares.

Não é possível quantificar quanto viés havia pois ninguém sabia ao certo o valor justo em cada caso, mas mesmo sem saber o alvo, estava claro que a variabilidade era o problema. Quem lida com uma organização espera um sistema que produza julgamentos consistentes, não ruído de sistema.

Uma característica definidora do ruído do sistema é que ele é indesejado, mas nem sempre a variabilidade em julgamentos é indesejada. No caso de críticos de cinema ela é bem-vinda. A diversidade de gostos ajuda a explicar o erro quando as preferências pessoais do profissional influenciam suas escolhas profissionais.

No caso de um analista econômico que alerta sobre uma recessão que ninguém havia previsto, ficaria

famoso por se afastar do consenso. Nos mercados, a seleção não funciona sem variação. Ruído de sistema é um problema dos sistemas e não do mercado. Quando investidores fazem avaliações diferentes sobre o valor de uma ação, uns ganharão dinheiro e outros não. Mas se um deles é solicitado para avaliar o valor de uma ação em nome de uma firma e descobrimos que outros investidores fizeram avaliações muito diferentes, a firma apresenta um ruído de sistema, e isso é um problema.

Uma auditoria de ruído foi realizada em uma firma de gestão de ativos, onde 42 investidores experientes estimaram o valor justo de uma ação (*valor que para eles seria indiferente comprar ou vender*), baseando sua análise nos lucros e prejuízos simplificados, balanço patrimonial, demonstrativos do fluxo de caixa para os 3 anos anteriores e previsões para os dois anos seguintes. O resultado foi uma variabilidade média de 41%.

Uma concepção errada sobre a variabilidade indesejada nos julgamentos é que não são significantes pois os erros aleatórios se cancelam. Se uma apólice de seguros está acima do preço e outra abaixo, a companhia cometeu dois erros. Em sistemas ruidosos os erros se somam.

Esses ruídos permeiam o universo das organizações e ninguém se da conta. Esse fenômeno é chamado de ilusão de concordância. Nós achamos que todos veem o mundo da mesma forma que nós. Temos poucas oportunidades de notar que as regras consensuais são

vagas, suficientes para eliminar algumas possibilidades, mas não para especificar uma resposta positiva compartilhada para um caso particular.

Uma corretora entrevistada descreve sua evolução em seu departamento. Ela narra que no início de sua carreira discutia 75% dos casos com seu supervisor, com os passar do tempo foi ganhando confiança em seu julgamento e foi tida como autoridade no assunto. Ela desenvolveu confiança no seu julgamento, exercendo-o antes de mais nada. A confiança é alimentada pela experiência dos julgamentos feitos com fluência e facilidade.

Para a corretora a ilusão de concordância só se desfez com a auditoria de ruído. A discordância traz desconforto, é preferível evitar o conflito.

## **DECISÕES SINGULARES**

Os ruídos em decisões recorrentes podem ser demonstrado por uma auditoria. É simples medir a variabilidade indesejada observando outros profissionais tomarem decisões em casos similares. Os casos singulares são mais complexos pois não são julgamentos feitos de forma recorrente pelo indivíduo ou grupo, não possuem estratégia pronta, são únicos.

Em 2014 o vírus ebola se espalhava para a África Oriental, projeções estimavam que a doença se alastraria pelo mundo. Nesse momento o presidente Obama teve que tomar uma decisão única. Fechar ou não as fronteiras. Ele decidiu não fechar e mandar um contingente de 3 mil pessoas para ajudar a combater o surto. Essa é um caso de julgamento singular.

Definimos o ruído como uma variabilidade indesejada nos julgamentos sobre um mesmo problema. Como problemas singulares nunca se repetem, a definição não se aplica para eles. Não poderemos comparar a decisão do Obama em enviar profissionais de saúde para a África Ocidental ao invés de fechar as fronteiras com a decisão que outros presidentes americanos tomariam na mesma situação.

Não há um modo direto de observar a presença de ruído em situações singulares mas sabemos que está lá.

Se os fatores que influenciam a decisão fossem outros, a decisão poderia ter sido diferente.

Uma decisão singular deve ser vista como uma decisão recorrente tomada uma única vez, portanto as estratégias que reduzem o ruído nas decisões recorrentes melhoram a qualidade nas decisões singulares.

## **Sua Mente é um Instrumento de Medição**

Medição é o ato de usar um instrumento para atribuir valor a um objeto ou evento em uma escala. O julgamento pode ser descrito como uma medição em que o instrumento é a mente humana. A precisão é a finalidade da medição, aproximar-se da verdade, minimizar o erro. Julgamento é uma conclusão. Apesar do objetivo da medição ser a precisão, ela jamais é alcançada, sempre há viés e ruído.

Em estatística a medida usada para medir variabilidade é o desvio-padrão vamos utilizá-lo para medir ruídos. O ruído em seus julgamentos implica erro. Se o objetivo do julgamento é um valor real, dois julgamentos diferentes não podem estar certos. Algumas pessoas cometem mais erros que outras, seja por falta de habilidade ou treinamento. Precisamos compreender esses erros e medi-los.

## QUESTÕES DE JULGAMENTO

Esse livro é sobre julgamentos profissionais. A pessoa que os emite tem o objetivo de chegar a uma conclusão correta. Porém o próprio conceito de julgamento, já admite a incerteza de estar correto.

Julgamentos implicam julgamento pessoal, comportam incertezas sobre a resposta e admitem a possibilidade de que pessoas coerentes e competentes discordem. Existe um limite admissível para a discordância. Ela é mais fácil de ser aceita quando o julgamento é absurdo.

O julgamento opera da seguinte maneira: da descrição feita do caso algumas informações te chamam atenção e outras você deixa passar, informalmente integra as informações obtidas anteriormente, depois você converte as informações integradas em um resultado. As três etapas são fontes de variabilidade.

Um julgamento de probabilidade isolado nunca pode ser confirmado ou desmentido. Um julgamento inverificável, como são muitos julgamentos profissionais, são condicionais. A previsão é importante mas nunca poderá ser testada ou pode ser de longo prazo demais para conseguirmos cobrar dos profissionais uma resposta.

A verificabilidade não muda a experiência do

juízo. A impressão de que seu juízo está correto é um sinal interno de conclusão de juízo. Uma característica desse sinal é a sensação de coerência que faz parte da experiência de julgar. Ele não depende de um resultado real, e acontece em julgamentos inverificáveis e verificáveis.

A verificabilidade não altera a experiência do juízo, mas altera sua avaliação após o fato. Juízos verificáveis podem ser obtidos observando a medida simples da diferença entre o juízo e o resultado. Por exemplo: a previsão meteorológica afirmou que hoje faria 21 graus e fez 18 graus. Essa abordagem não funciona para julgamentos inverificáveis. Porém há outra maneira de avaliar um juízo, é observando o processo de juízo aplicado a um grande número de casos. Um comentarista por exemplo considera 70% a chance de 100 candidatos ganharem uma eleição, depois que o resultado é divulgado ele avalia o sucesso do juízo comparando o resultado com a previsão feita. Os julgamentos são verificáveis no conjunto.

Outra maneira de avaliar o juízo é focando na qualidade do processo que o levou a ele ao invés de só olhar para o resultado. Um analista qualificado frequentemente errará o número de alguma previsão feita da mesma maneira que um chimpanzé em um caso isolado pode acertar. Focar no processo em um mundo que cobra resultado é muito difícil.

Julgamentos preditivos são julgamentos que fazem previsões do resultado, julgamentos avaliativos buscam equiparar prós e contras entre várias opções. Uma decisão exige julgamentos preditivos e avaliativos.

Para medir o ruído precisamos apenas de múltiplos julgamentos sobre o mesmo problema, não precisamos conhecer um valor real. A diferença entre viés e ruído é essencial para aprimorar o julgamento, mesmo não conseguindo verificar se o julgamento está correto, conseguimos melhorá-lo medindo o ruído através das auditorias. Uma vez medido, em geral é possível reduzi-lo.

## **MEDINDO O ERRO**

Um viés consistente pode gerar erros custosos, se um gerente empolgado prevê que uma obra levará a metade do tempo que acaba levando por exemplo, pode trazer muito prejuízo. Assim como o ruído, se o gerente faz as previsões erradas, mesmo que se aproximem da média, os erros se somam trazendo prejuízos.

Em julgamentos profissionais sempre que a meta for precisão, viés e ruído desempenham o mesmo papel no cálculo do erro total. Por isso a medição e redução do ruído deve receber tanta prioridade quanto a medição e redução do viés. O desvio padrão pode ser usado para medir o ruído.

Podemos reduzir os erros quando não sabemos se nossas previsões estão certas? Se há um grande erro médio nas previsões, o melhor caminho é corrigi-lo, antes de tomar atitudes para melhorar a previsão, é necessário saber se está certa. Para saber o erro é só diminuir a previsão pelo resultado.

Após saber o resultado do evento previsto, conhecemos o erro de cada previsor. Mesmo depois do resultado observamos que o desvio padrão permanece o mesmo. Saber o verdadeiro resultado não acrescenta nenhuma informação sobre ruído nas previsões. O que passamos a conhecer é o viés, que consiste na média

dos erros.

O método dos mínimos quadrados, é uma regra proposta por Gauss para pontuar a contribuição dos erros individuais sobre o erro total. A melhor estimativa para ser usada é a que minimiza o erro total das medições disponíveis. Na nossa intuição a média é a melhor estimativa. Nesse método, elevamos os erros ao quadrado dando aos grandes erros um poder maior que aos pequenos.

Os papéis do viés e do ruído no erro são facilmente resumidos nas equações de erro. O erro ruidoso é positivo quando ele é maior que o viés e negativo quando é menor. A média dos erros ruidosos é zero.

### **Erro em uma única medição = viés + erro ruidoso**

O erro total é igual a soma dos quadrados do viés e do ruído. Sendo o ruído, o desvio padrão das medições que é igual ao desvio-padrão dos erros ruidosos.

### **Erro Total= Viés\*2 + Ruído\*2**

Viés e ruído desempenham papéis idênticos na equação do erro. São independentes entre si e recebem o mesmo peso no erro total, a diminuição do erro total será a mesma se o viés ou o ruído reduzirem.

A equação de erro e as conclusões que extraímos dela dependem do uso do EQM como medida do

erro total. Ela é aplicável em julgamentos puramente preditivos, como previsões e estimativas, que tenham como objetivo aproximar-se do valor real com máxima precisão (*o mínimo viés*) e máxima exatidão (*o mínimo ruído*). Ela não se aplica para julgamentos avaliativos porque o conceito de erro depende de um valor real.

## ANÁLISE DO RUÍDO

Quando focamos em um único caso, toda variabilidade de julgamento é erro e os dois componentes do erro são viés e ruído. Grande parte da variabilidade nos julgamentos de diferentes casos é intencional, juízes federais teriam pouca utilidade se dessem o mesmo parecer para todos os casos. Entretanto a variabilidade de julgamentos (*ruído*) no mesmo caso continua indesejável.

Analisaremos uma auditoria de ruído realizada com sentenças de juízes federais a fim de determinar a disparidade das sentenças de um modo mais sistemático. Não há modo de determinar o valor real de uma sentença para um caso particular. Tratamos a média de 208 sentenças para cada caso coletadas na pesquisa como se fosse a “*sentença justa*” para ele. Sabemos que na realidade esse pressuposto não está correto, essa média está enviesada. A variância de vieses no conjunto dos casos é uma importante fonte de erro e parcialidade. Esse livro é sobre ruído mas não podemos deixar que isso diminuía a importância de medir e combater o viés.

Em um mundo ideal, sem ruído e viés, todos as sentenças dos juízes seriam idênticas e a variação seria somente para diferentes casos. Essa não é a realidade, considerando a sentença média, alguns casos tem menos tempo de prisão e outros mais, essa variabilidade

que não é a mesma em todas as sentenças é o ruído.

A medida do ruído em cada caso é o desvio-padrão dos tempos de prisão atribuídos a eles. Para a média dos casos a sentença média foi de 7 anos e o desvio padrão de 3,4 anos. Se calcularmos a diferença média absoluta que consiste na média da diferença entre a sentença de todos os juízes para todos os casos obtemos o resultado de 3,8 anos. Uma diferença média absoluta de 3,8 anos entre juízes quando a média das sentenças é 7 anos é um resultado perturbador, inaceitável. Suspeitamos ainda que a quantidade de ruído enfrentada pelos réus em um tribunal de verdade é ainda maior já que possuem muito mais informações sobre os casos e quanto mais informação maior é a possibilidade de divergência entre os juízes.

Dividiremos o ruído em vários componentes, falaremos agora sobre os erros de nível. Erro aqui é a definição do desvio da média. A variabilidade dos erros de nível será encontrada em qualquer tarefa do julgamento. Essa diferença é atribuída às características do juiz: formação, experiência de vida, opiniões políticas... Dentre 9 sentenças de um juiz, a sentença média é a medida do seu grau de severidade. Esse número não tem nada a ver com justiça, diz respeito a reputação de austeridade dele.

O resultado depende do objetivo dos juízes: se é a neutralização (*remover o criminoso da sociedade*), reabilitação ou dissuasão. Juízes cujo principal objetivo

é a reabilitação tendem a dar prisões mais curtas e maior tempo supervisionado, por exemplo.

Ruído de nível é a variabilidade na média dos julgamentos que é igual à variabilidade dos erros de nível.

Nesse exemplo o ruído de sistema é 3,9 e o ruído de nível 2,4 anos. Iremos agora analisar o ruído de padrão que corresponde a diferença entre esses dois ruídos. Os juízes não são igualmente severos em seu veredicto em todos os casos, são mais austeros que sua média pessoal em uns e mais lenientes em outros. Esses são os desvios de padrão. O termo estatístico para o desvio de padrão é juiz x interação de caso que é o mesmo que juiz por caso. Parte das reações idiossincráticas aos casos talvez corresponda à filosofia pessoal do juiz na aplicação das sentenças, outras partes resultam das associações das quais ele mal tem consciência, um réu que o lembra do seu filho, por exemplo. Esses padrões não são mero acaso, eles podem voltar a ocorrer se o juiz acompanhar o mesmo tipo de caso outra vez. Ele acrescenta mais incerteza à loteria das sentenças.

### **Equação do ruído**

$$\text{Ruído de Sistema}^2 = \text{Ruído de Nível}^2 + \text{Ruído de Padrão}^2$$

Os juízes não teriam determinado as mesmas sentenças se tivessem julgado os crimes outra vez. Se o

juiz está de bom humor, talvez julgue com mais leniência que o normal. Essa variabilidade intrapessoal é conceitualmente distinta das diferenças intrapessoais estáveis que acabamos de discutir, o nome para a variabilidade que se deve a efeitos transitórios é ruído de ocasião.

## **RUÍDO DE OCASIÃO**

Assim como um jogador de basquete nunca arremessa a bola da mesma maneira duas vezes, nem sempre elaboramos julgamentos idênticos quando confrontados com os mesmos fatos em ocasiões diferentes. Depende do momento em que o profissional elabora o julgamento, seu estado de espírito, a sequência de casos mais fresca em sua mente e outros incontáveis aspectos da ocasião. O ruído de ocasião é a variabilidade observada nos julgamentos sobre um mesmo problema por uma mesma pessoa e afeta todos os julgamentos.

O ruído de ocasião não é fácil de ser medido. Ao formar uma opinião profissional a pessoa associa a ela os motivos que justifiquem seu ponto de vista, se apresentarmos o mesmo caso pela segunda vez, para diminuir o esforço e manter a consistência ela reproduz a resposta anterior.

É possível confirmar a existência de ruído de ocasião por meio de decisões profissionais registradas em big data e econometria. Efeitos estatisticamente significativos de tais fatores irrelevantes para o julgamento são evidências de ruídos de ocasião.

Descobrir todas as fontes extrínsecas do ruído de ocasião é impossível, mas as que podem ser encontradas ilustram a grande variedade dessas fontes. Se

esperamos controlar o ruído de ocasião, precisamos compreender os mecanismos que os produzem.

O efeito sabedoria das multidões indica que a média dos julgamentos independentes de diferentes pessoas melhora a precisão de um modo geral. Francis Galton em 1907 pediu que centenas de pessoas respondessem quanto pesava um boi, ninguém acertou. Por outro lado, a média de todas as respostas errou por apenas meio quilo. Embora suas estimativas individuais fossem muito ruidosas, eram livres de viés.

Vul e Pashler inspirados nesse fenômeno pediram a várias pessoas que respondessem a mesma pergunta duas vezes, sem ter conhecimento na primeira ocasião que voltariam a ela. A média das duas respostas eram mais precisas que ambas. A média de diversos julgamentos independentes produz um novo julgamento que é menos ruidoso, embora não menos enviesado que os julgamentos individuais.

O mesmo efeito se estende ao ruído de ocasião, combinando duas conjunturas de uma mesma pessoa podemos chegar mais perto da verdade. Porém se consultarmos uma segunda opinião nos aproximamos ainda mais da verdade.

O bootstrapping dialético, é um método que sugere que o indivíduo considere a situação sob um novo prisma, reavalie, antes de dar a segunda opinião sobre o fato. O ganho em precisão foi muito maior do que a

resposta imediatamente consecutiva. As respostas fornecidas pelo participante são amostras tiradas de uma distribuição interna de probabilidades, não selecionadas de forma determinista com base no conhecimento total que o indivíduo possui.

O indivíduo não é o mesmo o tempo todo, o humor tem uma influência mensurável no que você pensa, o que você puxa da memória, como interpreta sinais, o que nota no seu ambiente...o modo como julgamos pode depender de como nos sentimos. O estresse, cansaço e o clima também são fatores extrínsecos que afetam nosso julgamento induzindo ruídos de ocasião. O desempenho no mercado de ações é afetado por um dia ensolarado, por exemplo.

A falácia do jogador é a tendência em mudar de direção após tomar uma sequência de decisões na mesma direção. Essa também é uma fonte de ruído de ocasião.

O ruído de ocasião perto dos outros tipos de ruído é um dos menos expressivos na totalidade do ruído do sistema.

A variabilidade intrínseca do funcionamento cerebral afeta a qualidade dos nossos julgamentos e muito mais do que os fatores extrínsecos. Portanto o ruído de ocasião não pode ser totalmente eliminado.

## COMO O GRUPO AMPLIFICA O RUÍDO

A tomada de decisão em grupo é muito ruidosa, um grupo pode ir para todas as direções dependendo de fatores aparentemente irrelevantes. Quem fala primeiro, quem fala com confiança, quem sorri, quem gesticula no momento certo... todos esses fatores alteram o resultado.

Falamos anteriormente que agregar julgamentos de múltiplos indivíduos reduz o ruído, porém existem multidões sábias e também multidões que seguem tiranos, vivem sobre influência de uma ilusão compartilhada. Devido a dinâmica nesse grupo o nível do ruído pode ser elevado.

Um experimento foi feito para identificar uma causa particular de ruído: a influência cultural. Grupos de pessoas deveriam selecionar músicas em uma lista aleatória. As músicas tinham nomes evocativos. Alguns grupos podiam ver as escolhas das outras pessoas do seu grupo. Se as músicas boas ficassem sempre nos primeiros lugares e as ruins nos últimos, todos os rankings seriam idênticos. O resultado foi que as paradas musicais divergiram radicalmente. O nível de sucesso na condição de influência social era mais imprevisível que na condição independente. Os resultados do grupo podem ser manipulados facilmente porque a popularidade se auto-reforça.

Outro experimento foi feito, invertendo o ranking de

popularidade das músicas, as músicas impopulares apareciam como as mais populares e vice versa. O resultado foi que o ranking influenciou os participantes a elegerem como favoritas as músicas que na realidade não eram populares. Um crescimento inicial de popularidade se auto reforça enquanto uma proposta que atrai pouco apoio no início fica condenada. Na política também o fenômeno se confirma, observaram que democratas endossavam uma opinião dada por democratas e rejeitavam uma opinião quando vinha de republicanos, independente do teor da afirmação.

O efeito de um único voto positivo inicial é uma receita para o ruído. Seja qual for o motivo desse voto, pode gerar um desvio em larga escala na popularidade total. Grupos similares podem emitir julgamentos muito diferentes, enquanto grupos isolados podem emitir julgamentos aleatórios em uma nuvem de possibilidades.

A sabedoria das multidões prega que ao propor uma pergunta a um grupo grande, há muita chance de que a resposta média seja próxima ao correto. Agregar julgamentos pode ser uma ótima forma de reduzir viés e o erro. Mas quando todo mundo escuta todo mundo, as pessoas pode confiar no que os outros pensam ao invés de emitir o seu próprio julgamento. Se todos compartilhassem seu conhecimento, a deliberação em grupo poderia ser positiva.

As cascatas informacionais tornam o ruído nos

grupos presentes e até prováveis. Não é por irracionalidade que você segue a opinião dos outros, se você não tem certeza sobre quem escolher acaba por seguir os outros. As cascatas podem conduzir o grupo a condições horríveis ou o grupo pode já estar inserido em um movimento de cascata.

As pressões sociais também impactam as escolhas em grupo, alguém pode se calar para evitar ser grosso, idiota. Queremos mostrar que sabemos trabalhar em equipe, por isso seguimos a opinião dos outros. As cascatas sociais tem o mesmo efeito das informacionais.

Outra fonte de ruído é a polarização em grupo, quando conversam entre si as pessoas terminam em um ponto mais extremo alinhado as suas inclinações originais.

## **O ruído no julgamento preditivo**

Muitos julgamentos que fazemos são previsões, como são verificáveis podem ser avaliados, o que torna possível aprender sobre ruído e viés através deles.

Compararemos a seguir a precisão de previsões feitas por profissionais, máquinas e regras simples. Os profissionais ficam em terceiro lugar. Para chegar a essa conclusão é preciso avaliar a qualidade das previsões, utilizando uma medida precisa de previsão. Qual é o grau de covariância (*medida de variação entre duas*

*variáveis aleatórias*) entre as previsões? Por exemplo se o RH avalia o potencial de contratados, podemos no futuro avaliar o desempenho e ver o grau de covariância entre o potencial e o desempenho.

Potencial de Concordância é uma medida intuitiva da covariância que confirma por exemplo se os funcionários cujo potencial foi considerado alto na contratação também receberam qualificações elevadas no seu trabalho. A medida padrão usada pelos cientistas sociais é o coeficiente de correlação que varia entre 0 e 1 quando duas variáveis estão positivamente relacionadas. Nessa medida, a correlação entre duas variáveis é sua porcentagem de determinantes compartilhadas.

## JULGAMENTOS E MODELOS

A previsão de desempenho é uma ótima maneira para exemplificar o julgamento profissional já que muitas pessoas tem interesse em prever o desempenho pessoal no trabalho.

Usando como exemplo a pontuação de previsão de desempenho de duas candidatas nos quesitos: comunicação, liderança, habilidades interpessoais e técnicas e motivação elaboraremos um julgamento. Se considerarmos a informação, fizemos um cálculo rápido, consultarmos nossa intuição e elaborarmos um julgamento de forma informal teremos um “*julgamento clínico*”.

Ao participarem de um estudo considerando esses dados, psicólogos com doutorado contratados para fazer essa previsão alcançaram um grau insatisfatório de correlação ( $0,15$ ) com as avaliações de desempenho (*porcentual de concordância 55%*). Quando consideravam um candidato melhor que o outro a probabilidade que ele terminasse com uma avaliação de desempenho mais alta era de 55%.

As informações oferecidas eram satisfatórias para o sucesso da previsão? Como as informações poderiam ser combinadas para obter a correlação mais elevada possível com o desempenho?

Regressão múltipla é um tipo de modelo simples de previsão mecânica. Até que ponto o julgamento humano é eficaz em relação a uma fórmula? No modelo de previsão mecânica, a mesma regra é aplicada a todos os casos, cada variável preditora tem um peso que não varia de um caso para o outro.

O modelo simples apresenta limitações: o aumento de uma variável preditora produz sempre o mesmo efeito. Em uma intuição clínica uma alta pontuação em um quesito pode influenciar o resultado geral de maneira desproporcional.

De ilusão da validade chamamos a ilusão que sentimos ao realizar uma previsão de acordo com nossas convicções, prever o que irá acontecer no futuro é muito incerto. Diversos estudos mostraram a superioridade do julgamento clínico, que além de mais assertiva é mais rápida e barata.

Os modelos estatísticos de julgamento não acrescentam nada às informações que contém, a única coisa que faz é subtrair e simplificar, esse modelo não apresenta ruídos, não capta influências do contexto, e de seu estado mental. Reduzir o ruído de seus julgamentos sempre vai representar incremento de sua previsão preditiva, embora reduza a sutileza. A complexidade e sutileza com que julgamos geralmente não acrescentam nada a modelos simples. As regras complexas que inventamos ao julgar, muitas vezes não são válidas e muitas vezes não são aplicáveis

em qualquer contexto.

O ruído prejudica o julgamento clínico. Nos julgamentos preditivos especialistas humanos são facilmente superados por fórmulas simples- modelos da realidade, modelos de um juiz, até modelos gerados de forma aleatória. Há a necessidade de usamos modelos livres de ruídos com regras e algoritmos.

## REGRAS SEM RUÍDO

Algoritmo é um processo ou série de regras a serem seguidas em cálculos ou outras operações de resolução de problemas, especialmente por um computador. A partir dessa definição outros tipos de julgamentos mecânicos como os modelos simples também são considerados algoritmos. Todos eles superam o julgamento humano e a principal razão é que elas são livres de ruídos.

Alguns tipos de julgamento mecânicos serão analisados a seguir:

Modelo linear impróprio: muito superiores a julgamentos clínicos e com a eficácia similar aos modelos de regressão (*atribuem o peso preciso a cada variável preditora*). Esse tipo de modelo, atribui pesos iguais a todas as variáveis preditoras. Suas vantagens são não serem suscetíveis a acidentes de amostragem, não exigirem dados que não dispomos nem cálculos complexos.

Os modelos frugais ou regras simples se baseiam no uso de poucas variáveis porém altamente preditivas no cálculo. Esses cálculos podem ser realizados à mão de tão simples que são.

A partir da tecnologia de Inteligência Artificial as previsões podem ser feitas utilizando uma quantidade

grande de variáveis, dispondo de uma série de dados, podendo assim trabalhar nos detalhes, com as exceções, encontrando padrões desconhecidos ou negligenciados melhorando as previsões. Quando esse tipo de previsão é bem sucedida, a vantagem sobre o julgamento humano não é apenas ausência de ruído mas também a capacidade de explorar muito mais informações.

Se os modelos mecânicos se mostraram tão superiores ao julgamento humano, porque temos tanta resistência ao adotá-los? Pesquisas mostram que psiquiatras, médicos, juízes, apresentam várias objeções sobre o uso desses modelos entre elas perder seu emprego para as máquinas, aversão a computadores, formação ruim, além de desistirem de usá-los se se depararem com um erro.

## IGNORÂNCIA OBJETIVA

Tomadores de decisão costumam dar ouvidos a sua intuição. Aqui definimos essa intuição como: *“um julgamento para dado curso de ação que vem à mente com uma aura de convicção de retidão ou plausibilidade, mas sem razões ou justificativas...”* Essa sensação de saber sem saber porque é o que chamamos de sinal interno de conclusão do julgamento. É uma experiência emocional gratificante, uma sensação agradável de coerência que nos faz sentir que o julgamento obtido está correto.

Pesquisas sugerem que tomadores de decisão preveem que vão acertar em seus julgamentos em 80% de porcentual de concordância enquanto o resultado das previsões demonstram que no mundo real a média dos julgamentos alcança 59%. Ninguém é capaz de prever esses eventos e circunstâncias. Tanto a incerteza intratável (*incertezas envolvidas no processo que não podem ser conhecidas*) quando a informações imperfeitas (*informações que poderiam ser conhecidas mas não são*) inviabilizam a previsão perfeita. Essas incógnitas não são problemas de viés ou ruído, são características da tarefa de julgar. Foi criado o termo ignorância objetiva para definir esses tipos de incógnitas.

A quantidade de ignorância objetiva vai depender de cada julgamento. Por exemplo, para muitas doenças,

médicos fazem previsões excelentes, para outras não. Mas onde quer que haja previsão há ignorância objetiva.

Philip Tetlock estudou previsões realizadas por quase trezentos especialistas sobre eventos políticos e descobriu que: *“as previsões de um especialista médio eram tão precisas quanto um chipanzé atirando dardos.”* Os especialistas analisam uma situação, descrevem um cenário convincente com grande confiança mas não sabiam de fato o que iria acontecer.

Os especialistas não tem culpa das falhas em suas previsões, eventos imprevisíveis fatalmente ocorrem e suas consequências são imprevisíveis. O limite de acerto em previsões não estão relacionados às capacidades cognitivas de quem realiza o julgamento e sim da ignorância objetiva. Eles merecem uma crítica por assumirem uma tarefa impossível e acreditarem que estão certos.

A ignorância objetiva aumenta para previsões futuras. Uma pesquisa mostrou que previsões de até um ano são difíceis de serem assertivas mas não impossíveis. Existem os super previsores que são os melhores nisso.

Modelos de julgamento mecânico demonstraram ser mais eficientes que os humanos mas a diferença não é tão grande. A inteligência artificial se sai melhor que os modelos mais simples mas seu desempenho está longe de ser perfeito.

A previsão perfeita não existe, pessoas que acreditam que são capazes de fazer previsões precisas estão negando o ruído e viés do julgamento, confundindo sua percepção subjetiva (*ignorância objetiva*) com um indicativo de validade preditiva, acreditando na previsibilidade de eventos que são imprevisíveis negando a realidade da incerteza. Essa atitude é a negação da ignorância. Isso acontece pois abrir mão da recompensa emocional da certeza intuitiva não é fácil. Por isso muitos tomadores de decisão rejeitam abordagens de julgamento que os privem de usar sua intuição. Enquanto os algorítmicos não forem perfeitos, o julgamento humano não será substituído.

## **O VALE DO NORMAL**

Agora que temos consciência da quantidade de ignorância objetiva que ocorrem nos julgamentos deveríamos sentir desconforto ao viver cercados de previsões falaciosas. Mas pelo contrário, continuamos pré-dispostos a fazer previsões ousadas para o futuro com base em poucas informações úteis.

Pesquisadores da universidade de Princeton, realizaram um estudo a fim de descobrir se os cientistas sociais seriam capazes de prever a trajetória de vida de famílias socialmente frágeis baseados nos dados disponíveis. O objetivo do estudo era medir o nível de ignorância objetiva. Os dados foram extraídos do Estudo de Famílias Frágeis e Bem- Estar Infantil, uma investigação de larga escala de indivíduos monitorado do nascimento aos quinze anos de idade.

Esse estudo se concentrou na previsibilidade de seis resultados observados na criança aos quinze anos, incluindo ocorrência de despejo recente, média escolar e uma medida geral de circunstâncias materiais da família. Convidaram equipes de pesquisadores para competir e gerar previsões precisas dos seis resultados escolhidos usando a massa de dados disponíveis sobre cada família. O resultado obtido pela equipe vencedora define o que há de recente no momento e passa a ser considerado ultrapassado na competição seguinte.

Os algoritmos de aprendizado de máquina tiveram um desempenho superior às previsões de modelos lineares simples e superariam juízes humanos. Já a Inteligência Artificial obtida sobre um modelo simples foi tênue e sua previsão preditiva permaneceu muito baixa. Ao prever despejos o melhor modelo teve uma correlação de 0,22 e percentual de covariância 57%, valores similares foram encontrados para resultados isolados. Dois dos seis resultados eram agregados que são muito mais previsíveis, o resultado das correlações preditivas da média escolar das crianças por exemplo foi de correlação 0,44 e percentual de covariância 65%. A principal conclusão foi que uma grande massa de informação preditiva não basta para prever eventos isolados na vida das pessoas - e até a previsão de agregados fica limitada.

Muitas correlações no campo das ciências sociais está nessa faixa. Correlações muito mais elevadas como 0,60 que preveem altura e tamanho de pé no adulto são comuns para medições físicas mas muito raras nas ciências sociais. A previsibilidade limitada de resultados isolados no desafio traz uma mensagem preocupante sobre a diferença entre compreensão e previsão. O estudo de Famílias Frágeis é considerado um tesouro das ciências sociais, os acadêmicos que o produziram esperavam que contribuiria para a compreensão de vida de famílias carentes mas infelizmente fazer previsões refinadas de eventos individuais não é possível.

No campo das ciências sociais e na maioria das

conversas do dia a dia, compreender alguma coisa significa compreender sua causa. Os sociólogos que coligiram e estudaram milhares de variáveis no Estudo das Famílias Frágeis buscavam as causas dos resultados que observaram. Compreender é descrever uma cadeia causal. A capacidade de fazer uma previsão é a forma de medir se tais cadeias causais foram identificadas. E a correlação (*medida de precisão preditiva*) é uma medida de quanta causação podemos explicar.

Correlação não implica em causação mas causação implica em correlação. Onde houver um elo causal, devemos encontrar correlação. Onde houver causalidade, deveríamos de ser capazes de fazer previsão e a correlação, a precisão dessa previsão, é uma medida de quanta causalidade compreendemos. Baseado nisso a conclusão dos pesquisadores de Princeton é a seguinte: a extensão em que os sociólogos conseguem prever eventos como despejos, tal como atestada pela correlação 0,22, é um indicativo de quanto compreendem (*ou de quão pouco compreendem*) as trajetórias de vida dessas famílias. A ignorância objetiva determina um teto não só para nossas previsões como também para nossa compreensão.

No pensamento estatístico você se pergunta o que leva por exemplo as famílias frágeis a despejos e outros resultados de vida. Sua preocupação com os conjuntos, a população de famílias frágeis e as estatísticas (*médias, variâncias, correlações*) que as descrevem. Já no raciocínio causal narrativas em que eventos, pessoas

e objetos específicos se afetam mutuamente, são produzidas. Por exemplo: casos de diversas famílias carentes são acompanhados. Uma família é despejada, mas você que está acompanhando o caso sabia que Jennifer Jones que é a provedora da família perdeu o emprego. Não conseguiu encontrar outro emprego, deixou de pagar o aluguel, implorou para o senhorio não despejar ele... O despejo dessa família é triste mas não surpreendente, pelo contrário, é o fim lógico de uma cadeia de eventos.

Quando cedemos a sensação de inevitabilidade, não percebemos como as coisas poderiam ter sido diferentes. As narrativas alternativas são tão pouco surpreendentes quanto a principal- se o fim é conhecido. Seja qual for o resultado, o raciocínio causal faz com que pareça totalmente explicável e até previsível.

A explicação lógica para essa observação é que alguns eventos são surpreendentes para nós como uma pandemia, mas a maioria deles é esperado como um menino de oito anos voltar para a casa depois da escola. No “vale do normal” os eventos se desenrolam como o desejo dos Jones, em retrospecto, parecendo normais embora não fossem esperados ou previstos. A busca da causa do despejo dos Jones cessa quando uma boa narrativa é encontrada. Isso aconteceria mesmo se o resultado fosse oposto. Muitos eventos em uma história normal são autoexplicativos.

Isso é o que entendemos sobre compreender uma

história e é o que faz a realidade parecer previsível - em retrospecto. Como o evento se explica em si temos a impressão de que poderia ter sido previsto. Dificilmente, um grande movimento no mercado de ações fica sem explicação. No momento que os investidores recebem uma explicação de uma queda nos índices ficam preocupados com a notícia se a explicação é sobre uma alta continuam otimistas. Rotulamos o resultado como surpreendente somente quando não podemos ajustar o nosso modelo de mundo para encontrar um resultado.

O modo causal de pensar sobre o mundo, poupa nosso raciocínio já que categorizamos os eventos em tempo real como normais ou anormais, já o estatístico exige treinamento especializado. Começa pelos conjuntos e considera casos individuais ocorrências de categorias muito mais amplas. O caso dos Jones não é visto como fruto de uma cadeia de eventos e sim como um resultado estatisticamente provável ou improvável. Basear-se no raciocínio causal é uma fonte de erros e no estatístico é uma maneira de evitar os erros.

## HEURÍSTICAS, VIESES E RUÍDO

As heurísticas são respostas simplificadoras produzidas no pensamento rápido e intuitivo quando são confrontadas com uma questão complicada. Normalmente elas produzem respostas adequadas mas por vezes levam a vieses que chamaremos de erros de julgamento sistemático. Múltiplos experimentos mostraram que pessoas dão respostas idênticas às duas questões. Mas similaridade e probabilidade são muito diferentes.

Imagine esse exemplo: Bill, tem 33 anos, é pouco criativo e na escola era bom em matemática e fraco em humanidades.

### **Para você, qual alternativa faz mais sentido?**

Bill se encaixa em minha ideia de uma pessoa que toca jazz como hobby.

Bill é se encaixa em minha ideia de um contador que toca jazz como hobby.

Bill tem mais o perfil de um contador que toca jazz como hobby do que uma pessoa que toca jazz como hobby. Agora considere o seguinte exemplo, qual é a mais provável?

Bill toca jazz como hobby.

Bill é contador e toca jazz como hobby.

Você pode ficar tentado a escolher a segunda mas a lógica é escolher a primeira. Se a informação que ele toca jazz está nas duas respostas, ela tem alta probabilidade de ser correta. Acrescentar detalhes a uma descrição só serve para torna-la mais provável. A questão da similaridade é mais fácil, o que tende a fazer dela a preferida quando é pedido que se avalie a probabilidade. A heurística para responder uma pergunta difícil é encontrar uma resposta mais fácil, a substituição de uma questão pela outra causa erros previsíveis chamados viés psicológicos.

O viés psicológico é percebido quando observamos que um fator que não deveria afetar o julgamento, tem um poder estatístico sobre ele enquanto um fator que deveria afetá-lo não tem.

O viés psicológico pode ter outra variação além da similaridade. Como estimar a probabilidade de um aluno passar em uma prova? As taxas de reprovações são altas mas com certeza o fator determinante seria o desempenho do aluno. Pensando assim, você negligência o que é chamado de taxa base. Ao utilizar a taxa de reprovação como ponto de partida para a previsão de probabilidade, você agrega essa informação relevante (*taxa base*) junto com as informações sobre o aluno.

Bill é contador e toca jazz como hobby.

Você pode ficar tentado a escolher a segunda mas a lógica é escolher a primeira. Se a informação que ele toca jazz está nas duas respostas, ela tem alta probabilidade de ser correta. Acrescentar detalhes a uma descrição só serve para torna-la mais provável. A questão da similaridade é mais fácil, o que tende a fazer dela a preferida quando é pedido que se avalie a probabilidade. A heurística para responder uma pergunta difícil é encontrar uma resposta mais fácil, a substituição de uma questão pela outra causa erros previsíveis chamados viés psicológicos.

O viés psicológico é percebido quando observamos que um fator que não deveria afetar o julgamento, tem um poder estatístico sobre ele enquanto um fator que deveria afetá-lo não tem.

O viés psicológico pode ter outra variação além da similaridade. Como estimar a probabilidade de um aluno passar em uma prova? As taxas de reprovações são altas mas com certeza o fator determinante seria o desempenho do aluno. Pensando assim, você negligência o que é chamado de taxa base. Ao utilizar a taxa de reprovação como ponto de partida para a previsão de probabilidade, você agrega essa informação relevante (*taxa base*) junto com as informações sobre o aluno.

Outro tipo de viés psicológico é a heurística da disponibilidade, ocorre quando você utiliza a impressão de casos que vem facilmente à mente no lugar de um julgamento sobre frequência. Quedas de avião por exemplo, sua percepção de risco cresce após um incidente ser noticiado em exaustão. Essa avaliação deveria se basear em uma média de longo prazo mas recebem mais peso pois nos veem à mente com mais facilidade.

Já o viés de conclusão ou prejulgamento ocorre quando iniciamos o processo de julgar inclinados a chegar a determinada conclusão. Deixamos nosso Sistema 1 (*sistema de pensamento rápido e intuitivo*) sugerir uma conclusão e utilizamos o Sistema 2 (*sistema de pensamento deliberativo*) para coletar informações para elaborar argumentos que apoiem nosso prejulgamento. A evidência selecionada é distorcida devido ao viés de confirmação e desejabilidade. Reunimos e interpretamos a evidência em favor do julgamento que queremos que seja correto.

Os prejulgamentos normalmente trazem um componente emocional. Chamamos de heurística afetiva, o processo de determinar o pensamento a partir dos próprios sentimentos.

Outro viés de conclusão é o efeito de ancoragem cuja influência de um número arbitrário cresce quando precisamos realizar um julgamento quantitativo.

A coerência excessiva é a formação de impressões

coerentes que são difíceis de mudar. A reação inicial positiva ou negativa afeta intensamente suas escolhas. Por exemplo, ao descrever uma pessoa com quatro características escritas em dois cartões, se no primeiro estiverem escritos dois adjetivos positivos e o segundo negativos, não mudará completamente a percepção sobre a pessoa pois ela já havia sido formada a partir dos adjetivos que foram relevados primeiro.

Os vieses psicológicos causam ruídos. Na substituição, se todo avaliador comete o mesmo erro não há ruído. Mas a resposta pode variar de pessoa para a pessoa. Os prejulgamentos geram viés e ruído, as diferenças individuais nos vieses podem causar ruído no sistema. A coerência excessiva pode gerar viés e ruído, contanto que a sequência de informações e o significado atribuído a ela sejam idênticos para todos.

Os vieses psicológicos com frequência produzem erros compartilhados, mas quando há grandes diferenças individuais nos vieses (*diferentes prejulgamentos*) ou quando o efeito dos vieses depende do contexto (*diferentes gatilhos*), haverá ruído.

## A OPERAÇÃO DE EQUIPARAÇÃO

Julgamento é uma operação que atribui valor escalar a uma impressão subjetiva. Equiparação é uma parte essencial dessa operação pois se trata de encontrar um valor na escala de julgamento que corresponda ao objeto ou fato julgado. Por exemplo: *“Atribua de 1 a 5 estrelas a sua experiência de compras desta manhã.”*

Alguns julgamentos são expressos por escalas qualitativas como diagnósticos médicos, outros quantitativas como volume, peso, temperatura. Nos julgamentos quantitativos podemos distinguir qual é maior, mais pesado, mais quente... O ser humano tem uma capacidade intuitiva de comparar intensidades em dimensões sem relação entre si, projetando uma escala de intensidade sobre a outra. Você pode equiparar a intensidade de seu apreço por diferentes compositores aos edifícios da sua cidade. As faixas de valores para uma escala dependem do contexto assim como o significado de palavras como grande e pequeno dependem inteiramente de um referencial.

No exemplo abaixo veremos como as previsões comparadas podem causar viés:

Julie lia fluentemente aos quatro anos de idade. Ela está cursando a faculdade. Dentro da escala padrão (*de 0,0 a 4,0*), defina sua média na faculdade.

Primeiro você avaliou a precocidade em que Julie começou a ler, essa avaliação foi fácil. Depois colocou essa precocidade dentro de uma escala, considerando que Julie começou a ler antes das outras crianças mas existem crianças que começaram aos 2 anos de idade, antes de Julie. Ela foi classificada entre crianças atipicamente precoces mas não extraordinariamente. Ao definir a média você sem perceber classificou também Julie entre as médias atípicas porém não extraordinárias fazendo uma previsão comparada, de forma rápida e descomplicada (*utilizando o sistema 1*).

A previsão intuitiva da média de Julie é um exemplo do mecanismo psicológico de substituição de uma pergunta difícil por uma fácil. A substituição acontece somente se a informação disponível for relevante, nesse caso o fato informado poderia ser interpretado como plausível de inteligência se tornando um substituto aceitável.

Substituir uma questão por outra causa erros quando as respostas são diferentes. Para perceber o porquê basta imaginar quantas coisas poderiam ter acontecido com Julie desde que tinha 4 anos de idade. Qualquer evento poderia ter afetado seu desempenho da faculdade. Uma previsão comparada se justifica apenas se a precocidade de leitura e média na faculdade estiverem perfeitamente correlacionadas. Por outro lado ignorar a informação sobre a idade da alfabetização de Julie também seria errado. A previsão ideal consideraria os dois extremos, o conhecimento perfeito

e o conhecimento zero.

Quando não temos nenhuma informação sobre um caso (*conhecimento zero*), a melhor estimativa é a média. Nesse caso a melhor previsão consideraria a média e a informação sobre a idade. O valor dependeria de qual das duas informações utilizaríamos como variável preditora (*a mais determinante*).

Embora os resultados possam ser absurdos é difícil resistir a previsões que equiparam evidências. O gerente de vendas geralmente presume que o melhor vendedor do último ano continuará a superar o desempenho do resto da equipe.

Como medida de correção para previsões intuitivas, sugerimos assumir a visão de fora. No caso de Julie, significa ancorar a previsão na média aritmética dos resultados.

Existe um limite humano para atribuir categorizações distintas aos estímulos em uma dimensão, ele gira em torno de sete denominações (*rótulos de valor*). Nossa capacidade de comparar valores entre dimensões de intensidade não pode ser melhor que a nossa capacidade de atribuir valor a essas dimensões. As pessoas podem ser treinadas para fazer distinções mais refinadas, como juízes discriminando graus de severidades, mas pra isso, as categorias devem existir de antemão e ser claras. Quando precisamos categorizar de forma intuitiva e rápida como por exemplo categorizar um conjunto de

linhas definindo em uma escala a mais longa, é mais difícil.

Para superar essa dificuldade em vez de categorizar é melhor comparar. As comparações explícitas entre os objetos do julgamento favorecem discriminações muito mais refinadas do que qualificar os objetos um a um. Se você faz uma idéia aproximada da riqueza das pessoas, é melhor comparar pares de indivíduos na mesma faixa do que categorizá-los individualmente.

Muitas tarefas do julgamento exigem equiparar casos individuais a uma categoria por escala ou usar uma série ordenada de adjetivos. Essa tipo de comparação é ruidoso porque é grosseiro. Os indivíduos podem diferir na interpretação das denominações até quando concordam sobre a substância do julgamento.

## ESCALAS

As escalas de respostas são fontes onipresentes de ruído. As pessoas podem diferir em seus julgamentos não porque discordem da substância mas porque usam a escala de maneira diferente. Ao classificar o desempenho de um funcionário, você pode dizer que em uma escala de 0 a 6, a avaliação dele é 4. O que no seu entendimento é muito bom. Outra pessoa poderia atribuir a nota 3 para o mesmo funcionário, e também considerar muito bom. A ambiguidade no fraseado das escalas é um problema geral.

Julgamentos expressos por frases também são fontes de ruído devido as interpretações de quem fala ou por quem escuta.

Os Estados Unidos permite a júris em casos civis impor reparações punitivas que suplementam as monetárias. Quando um produto causa um dano, por exemplo, além dos queixosos receberem uma indenização para saldar contas médias e eventuais prejuízos , pode ser atribuída uma reparação punitiva para mandar um recado a empresa acusada e outras semelhantes. A grande preocupação é o caráter imprevisível dessas punições. Por vezes são negadas, outras são surpreendentes e arbitrarias.

As determinações de punições justas são muito debatidas. Uma questão considerada difícil para os

estudiosos pode ser fácil para pessoas comuns que utilizam o método de substituição de questões difíceis por fáceis.

O principal motivador das punições é o quanto o julgador considera ultrajante a atitude em questão. A intensidade do ultraje determinará a intensidade da punição. Além desse fator motivador, deveremos levar em consideração o resultado da ação.. o dano, o prejuízo, a consequência sofrida pela vítima. Pesquisas demonstram que o dano causado à vítima não altera de maneira determinante o sentimento de ultraje, mas na hora de aplicar a penalização esse fator contribuiu para o aumento da punição assim como aumentou o valor monetário da indenização.

Outro fator importante a ser analisado é o quanto diferentes jurados concordam quanto ao grau de severidade na punição mas divergem amplamente ao traduzir seu intento punitivo em uma escala monetária.

Quando não se conhece o resultado de um julgamento, assume-se que a média entre os julgamentos, em um pesquisa por exemplo, é o valor justo da pena. Qualquer desvio desse valor é considerado um erro, um ruído. O ruído de sistema pode ser decomposto em ruídos de nível (*variabilidade no grau de severidade dos jurados*) e ruído de padrão (*variabilidade em como dado jurado responde a casos particulares*).

Comparando a incidência de ruído nos três momentos das reparações punitivas (*ultraje, punição do réu e valor da indenização financeira*) observamos que a decisão do quanto a atitude é ultrajante apresenta grande incidência de ruído. Avaliar um comportamento a partir de uma escala expressa a partir de expressões como: “*inteiramente aceitável, absolutamente ultrajante...*” é muito vago. A falta de clareza nas classificações torna um pouco de ruído inevitável. Já a definição da punição, mesmo que seja considerada “*severa*”, ela deverá se enquadrar no limite delimitado por lei.

A escala monetária foi a que apresentou mais ruído pois ao atribuímos o valor de uma punição, normalmente atribuímos um valor arbitrário. Outro juiz julgando o mesmo caso também poderia atribuir um valor arbitrário, porém diferente, e isso causaria ruído.

Somos muito mais sensíveis ao valor relativo das coisas que seu valor absoluto. Temos mais facilidade em atribuir valor a partir de escalas de razão. “*O vizinho rico perdeu metade de sua fortuna...*” o fato de entendermos essa sentença expressa esse fato. Uma escala de razão pode ficar amarrada em uma única que âncora (*módulo*) intermediária. Um número para atribuir números para outras intensidades em função dele.

Perceberam em experimentos que por mais que os números absolutos fossem diferentes e apresentassem ruído, os valores proporcionais eram semelhantes.

Ao definir uma indenização monetária, se podermos contar com uma âncora fixando um valor de comparação, o julgamento fica muito mais fácil. Se nos for apresentado um caso similar com o valor de indenização aplicado, basta comparar o quão ultrajante foi a atitude do réu desse caso com o caso em questão e estabelecer um valor comparado ao já fornecido.

Quando não temos acesso a esse valor de referência, nossa previsão é arbitrária. Se utilizarmos a primeira previsão como âncora para as próximas, diminuiremos o ruído padrão (*variabilidade em como dado jurado responde a casos particulares*) e como consequência o ruído. A fim de diminuir também o ruído de nível (*variabilidade entre as previsões dos juízes*) as previsões de diferentes juízes podem ser colocadas através de rankings.

Infelizmente na justiça o processo julgado por jurados no tribunal real é o primeiro e único que verão e os jurados não tem acesso a dados de processos anteriores afim de extrair uma âncora, presumindo que o senso de justiça dos jurados os levarão a uma punição correta, uma capacidade que os seres humanos não tem.

## **PADRÕES**

Ao realizar um julgamento ou uma previsão, se tivermos poucas informações disponíveis ou informações que apontem para a mesma direção, esse problema é fácil. Quando as evidências disponíveis pintam um retrato coerente, nosso rápido e intuitivo Sistema 1 de pensamento não tem dificuldade de dar sentido a elas. Problemas fáceis como esses são simples de serem resolvidos e todos concordam com a solução.

Ao realizar um julgamento ou uma previsão, se tivermos poucas informações disponíveis ou informações que apontem para a mesma direção, esse problema é fácil. Quando as evidências disponíveis pintam um retrato coerente, nosso rápido e intuitivo Sistema 1 de pensamento não tem dificuldade de dar sentido a elas. Problemas fáceis como esses são simples de serem resolvidos e todos concordam com a solução.

Para ficarmos confiantes com um julgamento, nossa narrativa deve ser coerente e não podem haver alternativas atraentes. A não ser que você seja um especialista que sabe porque sua explicação está correta e sabe porque as outras alternativas não estão.

Ruído de padrão é um erro no julgamento individual de um caso que não pode ser explicado pela soma dos efeitos separados do caso e do juiz. Um investidor

cauteloso por exemplo, que deixa a cautela de lado quando é apresentado aos planos de uma start-up empolgante. Os erros de padrão surgem a partir da combinação de fatores transitórios (*como o humor do juiz no dia do julgamento*) e permanentes (*o entusiasmo de um empregador em contratar um funcionário que estudou em determinada faculdade*).

As associações evocadas em nós ao nos deparamos com informações de um caso são idiossincráticas e imprevisíveis, mas tendem a ser estáticas. As informações do caso que chamaram sua atenção podem não impactar da mesma maneira outras pessoas, causando ruído. Na ausência de dados do resultado, a média seria a previsão correta e ao desviar da média, os julgamentos já são considerados errados e por isso ruidosos, mesmo que estejam precisos.

Traços de personalidades são compreendidos como variáveis preditoras de comportamentos reais. O bom senso sugere que embora o comportamento possa ser motivado pela personalidade, ele também é reforçado pelas situações. O comportamento depende desses dois fatores. Os padrões característicos de reação a situações permanecem razoavelmente estáveis ao longo da vida.

Diferença de níveis entre juízes correspondem às diferentes pontuações de traços de personalidade que representam uma média de comportamentos em múltiplas situações. Os casos são análogos às situações.

## AS FONTES DO RUÍDO

O ruído tem mais peso que o viés no erro total. Quando muitas pessoas elaboram julgamentos individuais sobre múltiplos casos e o fazem apenas uma vez, não temos como saber se o erro de padrão (*variabilidade de julgamento entre diferentes juízes*), é transitório ou estável. Embora a interpretação conservadora sobre o ruído de padrão seja considerá-lo um erro transitório, ou seja um ruído de ocasião.

Evidências nos levaram a perceber que os julgamentos ruidosos produzidos por diferentes pessoas são em larga medida determinados por algo que não é um viés do indivíduo, tampouco é transitório ou aleatório: as reações pessoais persistem de um indivíduo particular a uma multiplicidade de características, que determinam sua reação a casos específicos. Sugerimos que o ruído padrão estável é na verdade mais significativo do que os demais ruídos do sistema.

Inúmeros estudos corroboraram com essa descoberta. Os indivíduos de um modo geral não concordavam entre si mas seu julgamentos permaneceram bastante estáveis, essa “*consistência sem consenso*” nas palavras dos pesquisadores fornece evidência clara de ruído de padrão estável. Embora haja a presença de ruído de ocasião e ela seja preocupante, nada sugere que a variabilidade

intrapessoal seja maior que as diferenças interpessoais.

O ruído é antes de mais nada um subproduto de nossa individualidade.

A despeito da evidência de sua presença e dos múltiplos mecanismos que o produzem, o ruído raramente é mencionado como um fator preponderante no julgamento pois sempre encontramos razões para explicar os eventos. Julgamentos que produzem resultados satisfatórios são normais e raramente questionados, uma história de sucesso se justifica por si mesmo. Quando o resultados são anormais, ruins ou surpreendente bons, sentimos a necessidade de explicar. As respostas vem do erro, talento natural... Apostas feitas no passado viram atos de genialidade ou loucura depois que resultado é conhecido.

O viés psicológico é uma explicação causal legítima de um erro de julgamento se pudesse ter sido previsto de antemão, detectado a tempo real ou caso ofereça uma previsão para o futuro. Uma explicação causal que se aplique apenas para um evento, se configura como uma afirmação vazia. Essas explicações são muito usadas para resultados de negócios mas necessitam de relatos causais para dar sentido a elas. Vieses como o erro de atribuição (*tendência a atribuir culpa ou crédito ao agente por resultados explicados por sorte ou circunstância*) e retrospecto (*explica de forma simples, resultados de julgamentos que não poderiam ser antecipados*) são mais algumas formas usadas para explicar julgamentos ruins.

Nosso modo de pensar é causal. Atentamos de modo natural para o particular, acompanhando e criando relatos causalmente coerentes sobre casos individuais, em que os fracassos são com frequência atribuídos a erros, e os erros vieses. A facilidade com que julgamentos ruins podem ser explicados não deixa margem para o ruído em nossos relatos de erro.

## **JUÍZES MELHORES PARA JULGAMENTOS MELHORES**

Em qualquer julgamento, há pessoas que se saem melhor do que as outras. Mesmo em “*multidões sábias*” a resposta é melhor quando os indivíduos são capazes. Julgamentos são menos ruidosos e enviesados quando os juízes são bem treinados, inteligentes, experientes, dispostos a aprender com informações novas.

Muitos julgamentos não são verificáveis. Nesses casos a confiança no parecer desses profissionais baseia-se no respeito de seus pares. Respeito - especialista é como chamamos esse termo. Na ausência de valores reais para determinar quem está certo ou errado, as opiniões deles são as consideradas. Especialistas obtêm qualificações profissionais junto a comunidades profissionais e recebem treinamento e supervisão em suas organizações. Médicos em suas residências, advogados que aprendem com um sócio mais velho não aprendem somente técnicas do seu ofício, mas também são instruídos a seguir normas e métodos que dão a eles uma ideia de quais inputs devem ser levados em conta e de como elaborar e justificar seus veredictos.

Embora haja uma doutrina, sempre haverá ruído, porque ela dá margem à interpretações. Mesmo profissionais com treinamento idêntico, seguindo a mesma doutrina, divergirão em aplicá-la.

Além de conhecimento um bom julgador precisa de experiência. Os respeito-especialistas passam credibilidade devido a segurança no seu discurso, ajustando com facilidade os fatos que veem na construção de uma narrativas coerente.

Outro fator determinante para a qualidade dos julgamentos é a inteligência. Entende-se que os profissionais que se encontram na posição de julgar (*médicos, advogados, executivos*) são pessoas de capacidade intelectual avançada. Entretanto as diferenças dentro desse grupo, impactam os julgamentos. O teste de medição de inteligência GMA confirmou esse fato em estudos.

Por outro lado, pessoas de capacidade intelectual elevada, tendem a desenvolver melhor sua capacidade de convencimento, podendo criar alternativas que inspirem a credibilidade de seus pares, virando respeito-especialistas com a ausência de qualquer feedback da realidade.

Independente da capacidade mental, as pessoas diferem no seu estilo cognitivo. O teste de reflexo cognitivo (*TRC*) mede a frequência com que as pessoas rejeitam a primeira resposta (*reflexo impulsivo*) que vem a mente. Já o teste Competência de Tomara de Decisão Adulta mede a tendência das pessoas a cometer erros de julgamento típicos, como super confiança ou inconsistência nas percepções de risco.

O bom julgador busca ativamente informações novas capazes de contradizer suas crenças prévias, é metódico em integrar essas informações a sua perspectiva atual e está disposta e até ansiosa a mudar de ideia em consequência disso. Líderes firmes e decididos inspiram confiança mas podem não ser propensos a aceitar contra-argumentos.

## **DESENVIESAMENTO E HIGIENE DA DECISÃO**

Ex post e ex ante são duas maneiras de realizar um desenviesamento em julgamentos. Ex post é a maneira corretiva, realizada depois do julgamento, normalmente usada de forma intuitiva. O ex ante é a maneira preventiva de evitar vieses, o problema é que nem sempre se conhece de antemão a direção do erro, como exemplo dos vieses psicológicos, que são diferentes entre os juízes e consequentemente imprevisíveis.

O viés deveria ser detectado enquanto se passa por ele mas as pessoas raras mentes se dão conta disso. Essa cegueira em si é um viés chamado de viés do ponto cego. É mais fácil detectar os vieses de terceiros. Pensando nisso, uma alternativa é treinar um observador afim de que ele possa detectar em tempo real, os sinais de que um o mais vieses estão afetando decisões ou recomendações. Uma equipe de executivos em uma reunião de uma empresa por exemplo discutindo sobre uma decisão importante, enquanto um observador usa um checklist afim de diagnosticar se algum viés está desviando aquele julgamento da melhor conclusão possível.

Detectar o viés é inútil se os tomadores de decisão finais não estiverem decididos a combatê-los.

Essa alternativa traz bons resultados desde que os

observadores sejam bem escolhidos e não suscetíveis eles mesmos a sérios vieses. Esse papel pode ser ocupado por um supervisor, ou uma pessoa em cada uma das equipes de tomada de decisão para ser um caça viés ou um facilitador externo que tem a vantagem de ter uma perspectiva neutra.

Outro fator importante é a utilização de checklist adequado, nele não estarão escritos todos os vieses possíveis e sim os mais frequentes e importantes.

Os vieses são erros que normalmente conseguimos ver e até explicar, são direcionais já os ruídos são imprevisíveis, não podemos explicar nem reconhecer facilmente. As estratégias de redução de ruído são preventivas e tem o intuito de prevenir contra uma gama enorme de potenciais problemas. Chamamos isso de higiene de ruído. Fazer isso significa adotar técnicas redutoras de ruído sem conhecer os erros subjacentes que ela ajuda a evitar.

## SEQUENCIANDO INFORMAÇÕES NA CIÊNCIA FORENSE

A análise de impressões digitais são consideradas exatas, mas assim como os julgamentos, estão sujeitas a vieses psicológicos dos investigadores forenses. Esses vieses podem causar ruído.

Impressões latentes (*coletadas na cena do crime*) são borradas, parciais, irregulares, podem estar sobrepostas pelas próprias ou de outro sujeito, não tem a mesma qualidade e precisão da exemplar colhida em ambiente controlado e especializado. Portando a comparação entre elas exige a avaliação de um especialista, e é aí que pode haver o ruído.

O processo de comparação é composto pelas seguintes etapas: análise, comparação, avaliação e verificação. Depois de avaliada a impressão pode ser: identificada, excluída ou inconclusiva. Quando é identificada (*as impressões se originam da mesma pessoa*) uma segunda pessoa analisa.

Por muito tempo esse método foi considerado inquestionável. Nos raríssimos casos em que ocorriam erros, eram atribuídos a incompetência ou fraude. Em parte, esse procedimento demorou a ser questionado pela dificuldade em se chegar na verdade, não é fácil ter certeza sobre quem cometeu o crime.

O primeiro passo utilizado na auditoria de ruído para avaliação de impressão digital foi comparar as constâncias de julgamentos de um profissional. No caso deles, não é simples memorizar as digitais o que afasta a hipótese de refazer a avaliação por se lembrar do resultado. Nesse primeiro momento os investigadores não sabiam que seriam testados. No segundo passo, eles sabiam que seriam testados e tiveram seus julgamentos acompanhados por anos. Caso houvesse diferença entre os julgamentos, seria um ruído de ocasião.

Na segunda análise das mesmas impressões, eles recebiam informações sobre o contexto do caso, que trouxeram viés psicológico de confirmação (*viés de confirmação forense*), o que é preocupante pois mostra que os investigadores tomam a decisão baseados no contexto e não nas informações contidas na impressão. Essas informações mudam o que o examinador percebe e a interpretação dessa percepção. A segunda opinião que confirmará a sentença também pode ser induzida ao erro devido ao viés de confirmação da decisão do primeiro investigador. O espectro das cascatas do viés é o fenômeno em que um erro inicial motivado pelo viés de confirmação torna toda a informação enviesadas de maneira que os vieses influenciam o segundo, terceiro especialista.

Pesquisas concluíram que a má aplicação da ciência forense contribuiu para 45% de condenações injustas. É necessário saber a quantidade de cientistas forenses que cometem erros importantes. A maioria dos erros

leva os investigadores a pecar por excesso de cautela.

Para reduzir o ruído é necessário reconhecê-lo. Mas mesmo cientes do risco de viés os cientistas veem o julgamento ruído no julgamento dos colegas e não no próprio.

Para fazer uma higiene de decisão pra diminuição de ruído, nesses casos é necessário: sequenciar informações para limitar a formação de intuições prematuras dando aos investigadores somente as informações necessárias no momento necessário. Esse procedimento é chamado de desvelamento sequencial linear. Outra recomendação é documentar sua análise das impressões latentes antes das impressões exemplares para saber se as duas batem. Esse processo diminui o risco de que uma impressão inicial envie todo o processo.

Outra recomendação é que o examinador que irá fazer a verificação não tenha acesso ao parecer do primeiro examinador.

## SELEÇÃO E AGREGAÇÃO EM PREVISÕES

Muitos julgamentos envolvem previsões. Escolhas fundamentais de instituições privadas e públicas dependem dessas previsões. Analistas de previsões (*analizam quando e porque elas dão errado*) fazem uma nítida distinção entre viés e ruído. Previsores são em geral enviesados, tendem a ser super confiantes em suas previsões. Também são ruidosos, até entre especialistas, a inconfiabilidade é uma das principais fontes de erro, nem sempre concordam entre si (*ruído de ocasião*).

Um estratégia para redução é selecionar juízes melhores para realizar julgamentos melhores outra seria agregar múltiplas estimativas independentes (*uma estratégia de higiene de decisão*). O modo mais fácil de agregar previsões diversas é tirando sua média. Essa conta significa dividir pela raiz quadrada o número médio de julgamentos, como consequência, se tivermos 100 julgamentos reduziremos 90% o ruído. Essa lei estatística é o mecanismo por trás da sabedoria das multidões. Essa estratégia não altera em nada o viés, portanto seu efeito no erro total (*eqm*) depende das proporções de viés e ruído que ele contém. Por isso que a sabedoria das multidões funciona melhor quando os julgamentos são independentes, com menor probabilidade de viés. A média de múltiplas previsões aumenta em grau elevado a previsão, como em um “*consenso*” de analistas de investimento sobre

o mercado de ações.

A estratégia de multidão seleta, seleciona os melhores juízes segundo a precisão de seus julgamentos recentes e tira a média dos julgamentos de um pequeno número de juízes. É eficaz e mais fácil para o tomador de decisão que respeita o conhecimento especializado para compreender e adotar uma estratégia baseada na agregação e seleção.

Usar mercados preditivos é uma maneira de fazer previsões agregadas. O método Delphi também é uma opção, apesar de ser desafiadora. São múltiplas rodadas em que os participantes fazem estimativas anonimamente a um moderador, justificando sua escolha. Esse processo estimula a convergência de estimativas.

Para saber se os previsores são bons, suas estimativas de probabilidade tem que corresponder a realidade. As boas previsões devem ser atualizadas com bases em novas informações. Beta perpétuo é o grau de comprometimento da pessoa com a atualização da crença e o aperfeiçoamento. Além disso elas tem que ser bem calibradas (*fazer previsões probabilísticas que correspondam a realidade*) e ter alta resolução (*fazer previsões pouco distantes da realidade*). Os supervisores (*os melhores previsores*) são inteligentes mas além disso pensam em termos analíticos e probabilísticos. Ao invés de responder a partir da intuição, levantam uma série de questões secundárias e

tentam responder a elas. Além disso adotam a visão de fora (*neutra e imparcial*) e levam a taxa-base (*a raridade em que um evento semelhante ao em questão acontece*) muito a sério.

Essas três informações proporcionam melhora para o julgamento: treinamento dos juízes, a formação de equipe a fim de promover a agregação e a seleção do melhor grupo baseado em taxa de previsão.

De acordo com Ville Satopã e Ttlock Mellers, há três razões para previsores terem melhores resultados: habilidade em selecionar no ambiente e analisar dados relevantes para cada previsão, serem menos suscetíveis aos diferentes gatilhos que levam ao ruído, ter uma tendência a incorrer ao lado do valor real da previsão. Esse processo melhora os vieses psicológicos e por consequência os ruídos causados pela variação de julgamentos que esses vieses geravam.

Combinar julgamentos complementares dentro de um grupo selecionado de supervisores aumentará ainda mais a validade da precisão já que cada um está olhando de um ângulo e trazendo informações complementares. Para montar essa equipe primeiro o melhor juiz deve ser escolhido, em seguida um juiz que agregue alguma habilidade nova para o grupo, mesmo que os julgamentos difiram entre si, mas a média desse grupo ruidoso será mais precisa do que um grupo unânime desde que os julgamentos sejam independentes.

## **DIRETRIZES NA MEDICINA**

Na medicina existem diagnósticos que não envolvem julgamento. A saúde pública progride removendo o julgamento e partindo para o cálculo através de procedimentos rotineiros e mecânicos transformados em diretrizes diagnósticas, já que em seus julgamentos os médicos podem ser ruidosos e errar. Uma prática-padrão é obter uma segunda opinião, sempre que diverge da primeira, há ruído.

Na medicina, a magnitude dos ruídos é enorme e abordagens como higiene da decisão são utilizadas afim de reduzi-lo. Eles não se limitam aos diagnósticos, tratamentos também são muito ruidosos. Quando o ruído está presente um médico pode estar certo e o outro errado. A habilidade faz muita diferença, a variação de habilidade corresponde a 44% do ruído nas decisões diagnosticas. Portanto, políticas criadas para melhorar habilidades mostram melhor desempenho do que diretrizes de decisão uniformes.

Ruído interpessoal, ou confiabilidade entre avaliadores, é medido pela estatística kappa. Quanto mais elevado o kappa, menos ruído. O diagnóstico de uma doença grave como câncer pode depender de uma espécie de loteria, determinada pelo médico que por acaso atende o paciente. Médicos por exemplo, em 31% ocasiões discordaram em avaliações de angiografias sobre o bloqueio em mais de 70% de uma

importante artéria.

Há também ruídos de ocasião, radiologistas as vezes oferecem uma opinião diferente ao avaliar a mesma imagem , discordando de si mesmos. Um estudo que não tem relação com diagnósticos continha a informação de que os médicos tendiam a passar mais exames no horário da manhã. Uma explicação para isso, pode ser o cansaço. Para não deixar o paciente seguinte esperando ainda mais, os médicos deixa de conversar sobre medidas profiláticas.

Em um extremo o diagnóstico para alguns problemas e doenças é essencialmente mecânico e não dá margem para julgamento, do outro não é mecânico mas é inequívoco. Em outras ocasiões médicos especialistas asseguram um ruído mínimo, e outros dão margem a julgamento e os critérios relevantes para diagnóstico são tão amplos que o ruído será substancial e difícil de reduzir.

O treinamento que aumenta a habilidade, múltiplos julgamentos especializados, algoritmos, inteligência artificial, são algumas das estratégias para redução de ruído. A tendência é que no futuro, a profissão médica fique cada vez mais dependente de algoritmos pois prometem reduzir vieses e ruídos, poupando vidas e dinheiro. Nesse livro abordaremos diretrizes para o julgamento humano e como elas produzem bons resultados.

Diretrizes funcionam não por eliminarem a necessidade de julgamento, a discordância permanece possível em cada etapa e logo na conclusão final, mas conseguem reduzir o ruído porque decompõe uma decisão complexa em uma série de subjuízos mais fáceis sobre dimensões pré-definidas. Elas oferecem uma descrição clara de como avaliar cada dimensão, simplificando os julgamentos formulados, reduzindo o ruído.

Psiquiatria é um caso extremo de ruído, diagnosticando o mesmo paciente, usando os mesmos critérios, psiquiatras com frequência discordam entre si. Outros tendem a diagnosticar muitos pacientes com determinada mesma doença. Os ruídos dessa área vem de vários motivos: diversidade de diagnósticos, grande quantidade de linhas de pensamentos diferentes além da inadequação da nomenclatura. Tal problema incentivou a criação do DM (*manual de diagnóstico de transtornos mentais*). Esse manual possui critérios bem detalhados, com uma escala clara. O uso dos critérios diagnósticos aumentou a confiabilidade dos diagnósticos psiquiátricos. O ruído foi reduzido, porém a área continuava a apresentar ruídos significativos. O principal motivo para o sucesso limitado é que os critérios de diagnóstico de alguns transtornos ainda são vagos e difíceis de operacionalizar.

O caminho seria, esclarecer critérios diagnósticos, evitando padrões vagos; produzir definições de

referência dos sintomas e seus níveis; usar entrevistas estruturadas na consulta... Esses passos parecem promissores mas a fé nos sintomas subjetivos do paciente, a interpretação dos sintomas feita pelo médico e a inexistência de uma medida objetiva, plantam a semente da inconfiabilidade diagnóstica e nesse sentido os psiquiatras são altamente resistentes a tentativas de diminuição do ruído.

## **DEFININDO A ESCALA EM ANÁLISES DE DESEMPENHO**

Em um mundo ideal, avaliar desempenho de funcionários dentro de uma empresa, não envolveria julgamento: fatos objetivos bastariam para determinar como as pessoas estão se saindo. Focar em um indicador por exemplo, pode resultar em avaliações equivocadas e acarretar efeitos prejudiciais. Até as metas de desempenho quantificáveis, devem ser avaliadas em um contexto, os funcionários não podem ser avaliados por métricas de desempenho objetivas, o que prova que todas as avaliações realizadas envolvem julgamento.

Essas avaliações de desempenho são extremamente ruidosas. Estudos mostram que a variância atribuída ao desempenho da pessoa corresponde de 20% a 30% da variância total e o resto (*de 70% a 80%*) é ruído de sistema, e todos os componentes de ruídos estão presentes. O ruído de nível ocorre quando um avaliador costuma atribuir notas mais altas aos funcionários. Esse ruído pode significar que os avaliadores têm impressões diferentes pelos avaliados ou que os avaliadores apenas utilizam a escala de classificação diferente. Já a fonte de ruído padrão ocorre quando um avaliador possui um padrão exclusivo para avaliar determinada pessoa. E o ruído de ocasião pode ocorrer se antes de realizar a avaliação, o julgador bate de carro, por exemplo.

*“A relação entre avaliação de desempenho profissional e empenho profissional é incerta.”* Pensando nisso algumas organizações se mobiliaram para desenvolver estratégias de redução de ruído como a agregação. A avaliação 360 graus é um exemplo de classificação agregada que consiste na média de classificações feitas por diversos avaliadores (*colegas, subordinados*). Embora seja uma ferramenta útil, essa classificação apresenta problemas como o excesso de informação gerada que passa a fazer parte das descrições de cargo, feedbacks, além de onerar o trabalho do avaliador.

*“Quando quase todo mundo recebe a maior qualificação possível, é justo questionar a validade das classificações.”* Para diminuir o inchaço nas avaliações elas podem ser padronizadas através do ranking forçado. Os avaliadores são impedidos de dar a maior classificação possível além de terem que obedecer a uma distribuição pré determinada. Esse método assegura a franqueza nas classificações.

Há menos ruído em julgamentos que usam escalas relativas que absolutas. Quando o funcionário é classificado em uma escala absoluta precisa haver uma equiparação para encontrar a pontuação que mais se assemelha com sua impressão, já na escala relativa, o indivíduo é comparado a um grupo em uma dimensão específica. Depois de comparado a outros, o supervisor deve declarar em percentagem a posição do funcionário na população especificada.

Essa abordagem traz a vantagem de classificar todos os funcionários em dimensões de cada vez. Estruturar o julgamento em dimensões é uma estratégia para diminuir ruído. A estratégia também diminui o ruído de padrão já que temos menor tendência a sermos inconsistentes ao comparar duas pessoas em uma mesma equipe. E a ordenação elimina o ruído de nível pois na classificação relativa, mesmo que um julgador seja mais leniente que o outro somente a média das classificações será diferente, a média dos rankings não.

O ranking forçado pode gerar alguns problemas como a confusão entre desempenho relativo e absoluto já que as avaliações relativas só são adequadas se a organização se preocupa com desempenho relativo as que forçam um ranking relativo quando se pretende medir um nível absoluto de desempenho, não agem com lógica. O outro problema é presumir que a distribuição forçada das classificações reflete a distribuição dos reais desempenhos subjacentes, escolhendo de maneira aleatória membros do grupo não é possível garantir que representem de maneira real o resultado recolhido do todo. Suponhamos que a equipe de um avaliador seja composta de cinco pessoas cujos desempenhos são indistinguíveis, forçar uma distribuição de desempenho diferenciada sobre essa realidade indiferenciada não reduz o erro e sim aumenta.

Mesmo com todos os esforços voltados para a

melhora da medição de desempenho nas organizações os resultados tem sido desanimadores. Geram informações imprecisas, acabam por desmotivar tanto quanto motivar os funcionários.. Muitas empresas aboliram esse ritual, mas para as que ainda realizam o procedimento, existem algumas estratégias de redução de ruído. Utilizar a escala certa assegurando um referencial comum, é uma delas. O referencial comum é descrito de forma específica afim de servir como âncora.

Além do referencial comum o treinamento referencial é outro importante passo para eliminar o ruído assegurando a consistência dos avaliadores. Os avaliadores são treinados para reconhecer diferentes dimensões de desempenho através de vídeos de sinopses de desempenho, afim de comparar os vídeos com as avaliações verdadeiras. Cada sinopse define um ponto de ancoragem na escala de desempenho de maneira que o novo avaliado seja comparado aos casos âncora.

## ESTRUTURA EM CONTRATAÇÕES

A contratação de um funcionário depende do julgamento de um contratador. A entrevista padrão (*uma das estratégias de contratação*) será o objeto de estudo desse capítulo. Pesquisas sobre entrevistas de emprego avaliam o seu sucesso comparando-as às avaliações do candidato no cargo. Basta comparar a correlação entre as duas referências. Um problema se faz presente, a avaliação de sucesso de um candidato no cargo é avaliada por classificações feitas por um supervisor. É razoável usar os julgamentos feitos pelo mesmo supervisor que realizou a contratação para avaliar o sucesso do funcionário no cargo.

A correlação entre as duas referências variam em estudos entre 0,2 e 0,3, uma correlação muito boa para ciências sociais mas não tão boa para servir de base para decisões. Entrevistas de emprego servem para outros propósitos além de formular um julgamento sobre o candidato, como promover a empresa para futuros candidatos, construir uma relação com futuros colegas, mas na tarefa principal (*seleção de talentos*) não têm sido um sucesso.

As entrevistas profissionais geram erro devido a ignorância objetiva. O desempenho profissional depende de coisas imprevisíveis que são impossíveis de ser consideradas durante a entrevista como sua adaptação ao cargo ou como eventos pessoais influenciam seu

trabalho. Essas entrevistas também são repletas de vieses. De forma involuntária o contratador costuma privilegiar candidatos culturalmente parecidos com ele além da aparência física sempre exercer um imenso papel na escolha.

É uma área cheia de ruídos, diferentes entrevistadores agem de forma diferente ao avaliar um candidato provocadas pelo ruído padrão. As empresas já esperam essas diferenças por isso promovem opiniões agregadas, pedindo que vários entrevistadores avaliem o mesmo candidato. Normalmente opinião agregada é formada a partir de um consenso entre os supervisores o que também traz problemas. O ruído de ocasião também está presente nas entrevistas, a primeira impressão que o entrevistador tem do candidato pode conduzir a entrevista. Até o aperto de mão pode ser significativo.

Os entrevistadores buscam encontram coerência nas respostas dos entrevistado de maneira que suas impressões iniciais façam sentido, encontram padrões imaginários em dados aleatórios , *“veem formas nos contornos das nuvens”* para encontrar lógica em respostas sem sentido.

Entrevistas não devem ser a única fonte de informação, existem testes, referências e outros inputs. A combinação desses fatores constroem um julgamento. Deve ser usada a agregação clínica ou mecânica? A mecânica já se provou ser mais eficiente

mas a mais usada nas empresas é a clínica.

O agregação é uma estratégia usada nesse contexto, quase todas as empresas agregam julgamentos de múltiplos entrevistadores sobre o mesmo candidato. O Google as vezes sujeitava o mesmo candidato a 25 entrevistas. O estudo de Bock concluiu que reduzindo esse número para quatro, o resultado tinha a mesma validade. Lembrando que os julgamentos só funcionam se forem independentes. O Google também utilizou a estratégia de Estruturar Julgamentos Complexos , que são definidos através de 3 princípios: decomposição, independência e julgamento holístico protelado. Decompor nesse caso é criar uma descrição clara, detalhada e realística sobre o cargo assim como as competências necessárias para ocupa-lo. A independência exige que as informações sobre cada avaliação sejam coletadas de forma independente. O princípio do julgamento holístico protelado consiste na combinação entre julgamento mecânico e não mecânico. A coleta de dados, informações, inputs coletados e avaliados no processo seletivo compõe a parte mecânica, mas a decisão final é um julgamento realizado pelo comitê que leva todas as evidências em consideração e as pesa holisticamente.

As evidências mostram a superioridade do julgamento estruturado (*com entrevistas estruturadas*), eles também são menos onerosos já que as interações não são presenciais.

## **PROTOCOLO DE AVALIAÇÕES MEDIADORAS**

O protocolo de avaliações mediadoras incorpora a maioria das estratégias de decisão já abordadas, pode ser aplicada amplamente sempre que um planejamento ou uma opção exija considerar múltiplas dimensões.

Assim como em uma seleção estruturada, a decisão estratégica de adquirir uma nova empresa, pode ser realizada a partir do princípio do protocolo de avaliações mediadoras. Nesse caso o primeiro passo seria realizar uma descrição detalhada da empresa que poderá ser comprada, um checklist de características (*avaliações mediadoras*) ou atributos que uma empresa deve possuir, depois a diretoria discute essas avaliações em dimensões separadas, sequencialmente e de forma independente. Depois disso decide sobre aceitar ou rejeitar o acordo.

Usar uma abordagem estruturada, nos força a postergar o objetivo de chegar a uma conclusão até termos feito todas as avaliações. Caso contrário, buscamos uma justificativa para confirmar nossas primeiras impressões.

Afim de compor uma lista de avaliações mediadoras eficaz, é necessário esboçar uma lista abrangente de avaliações independentes sobre o negócio. Qualquer fato relevante precisa ter seu lugar. Eles precisam ser

independentes de maneira que um fato importante influencie de preferencia apenas uma decisão. O objetivo dos itens dessa lista não é nesse momento levar a decisão final e sim avaliar a empresa sobre aquela ótica.

Depois de concluído o checklist, ele deve ser avaliado. Nesse exemplo foram usados conceitos como taxa-base (*a porcentagem de empresas que atingiram os níveis esperados*), classe de referência (*grupo de negócios considerados suficientemente comparáveis*), as análises foram comparativas, utilizando julgamentos relativos. Para evitar o efeito halo (*quando uma impressão geral transmitida influencia a avaliação de uma dimensão específica*) deveriam ser nomeados funcionários capacitados, para cuidar de cada avaliação de forma separada e independente.

Na hora de tomar a decisão, depois de aplicar o protocolo de avaliações mediadoras, notou-se que as avaliações não eram perfeitas, eram discrepantes, do tipo que levantariam um rico debate. Essa é a beleza das avaliações independentes. Cada assunto separado para o debate seria analisado separadamente e de forma independente antes de dar início ao debate final.

O método estimar-conversar-estimar, utilizado na fase decisiva, funciona da seguinte maneira: para cada avaliação, um funcionário resumia rapidamente os fatos principais (*que a diretoria já havia lido*) em seguida, os membros da diretoria deveriam fazer uma classificação anônima para essa avaliação, feito isso

as classificações eram projetadas no telão sem dizer quem havia feito, gerando uma leitura imediata da opinião independente de cada diretor antes de iniciar uma discussão. Esse método reduz o risco de influência social e das cascatas informacionais. A discussão, carregada de fatos e argumentos gerava a necessidade de haver outra estimativa e assim funciona o ciclo: estima, conversa e estima. Depois de avaliados todos os temas , chegou a hora de usar a intuição e realizar o julgamento final.

## **OS CUSTOS DA REDUÇÃO DE RUÍDO**

Uma objeção à eliminação de ruído é que os passos necessários são dispendiosos demais. Essa preocupação é legítima mas com frequência não passa de um pretexto. Uma pergunta deve ser feita: quanta precisão mais seria obtida, até que ponto é importante ter mais precisão e quanto tempo e dinheiro seriam gastos na tentativa de reduzir o ruído? Alguma forma de higiene de decisão certamente valerá a pena. Mas a crença de que é caro reduzir o ruído nem sempre está errada.

Temos que comparar custos e benefícios, as auditorias de ruído são muito importantes pois desempenham essa função: revelam se o ruído produz níveis ultrajantes de injustiça, custos elevados ou ambos. Caso sejam, o preço de redução de ruído, dificilmente é um bom motivo para fazer uma tentativa.

Uma objeção diferente é que algumas tentativas de redução de ruído podem ser responsáveis por gerar níveis inaceitáveis de erro. Essas tentativas podem ser perversas, agravando o problema ao invés de solucionar; podem ser inúteis e não alterarem em nada ou podem por em risco valores importantes. Apesar disso devemos rejeitar a ideia de que situações humanas são altamente variadas e de que juízes bons lidam com variações – o que pode corresponder a tolerar o ruído ou, rejeitar algumas estratégias de redução dele.

A regra de três delitos, condena a pessoa a prisão perpétua caso cometa três crimes. Ela foi criada para reduzir o ruído mas em contrapartida o preço do combate ao ruído é caro demais. Algumas pessoas podem passar a vida inteira na prisão mesmo que não tenham cometido crimes violentos, ou mesmo que a reabilitação seja um alternativa para elas. Regras complexas podem ser mais precisas mas podem ser mais injustas. Mesmo que algumas estratégias de ruído garantam erros excessivos, não deveríamos aceitar níveis elevados de ruído e sim conceber estratégias de ruído melhores.

Os custos excessivamente elevados na redução de ruídos costumam acontecer no uso de algoritmos. O uso de algoritmos também pode ser relacionado a uma consequência discriminatória com base na raça, gênero ou contra grupos desfavorecidos. Tenha ou não sido produzido para isso, o algoritmo poderia utilizar variáveis preditoras correlacionadas a raça ou gênero como: altura e peso estão correlacionadas a gênero, e o lugar onde a pessoa cresceu pode estar relacionado a raça. A discriminação também pode vir de dados brutos: se um algoritmo é treinado em um conjunto de dados enviesado, também será enviesado. Algoritmos utilizados para antever crimes afim de melhorar a alocação de recursos policiais utilizam dados existentes sobre criminalidade, refletem o policiamento em certos bairros ou denúncias excessivas, perpetuando ou exacerbando a discriminação, quando tratamos de certos delitos.

Recomendamos uma avaliação cuidadosa para assegurar que os algoritmos desconsiderem inputs inadmissíveis e testar se discriminam de uma forma censurável. Nesse sentido, é mais fácil identificar esses erros no algoritmo que nos seres humanos.

Podemos criar algoritmos que se saiam melhor que juízes humano em uma combinação que traga precisão e redução de ruído, não discriminação e imparcialidade. Embora um algoritmo preditivo em um mundo incerto dificilmente seja perfeito, ele pode ser muito menos imperfeito que um julgamento humano, ruidoso e enviesado.

## DIGNIDADE

Muitas pessoas desejam uma audiência individualizada onde são examinadas as circunstâncias particulares de cada caso, isentas do que veem como tirania das regras (*peças tratadas como membros de uma massa sem rosto, indiferenciada e cegamente sujeitada a alguma penalidade*). As audiências individualizadas geram ruído.

A misericórdia por exemplo, é ruidosa por não ser limitada por regras. Mesmo que os tomadores de decisão tenham consciência desse fato, podem rejeitar estratégias de ruído, sobretudo regras rígidas por garantir à pessoa a sensação de que será ouvida e tratada com respeito.

Algumas estratégias de redução de ruído não causam essa objeção, quando 3 juízes analisam o mesmo caso por exemplo, o julgamento continua individualizado mas o ruído diminui, diretrizes também. Ao contrário das regras rígidas que eliminam a discricionariedade e podem levar as pessoas a alegar que o processo resultante ofende sua dignidade.

Um sistema limitado de regras pode eliminar o ruído, o que é bom, mas também paralisar as normas e valores existentes, o que não é bom. Uma das vantagens defendidas por algumas pessoas, do sistema ruidoso, é que ele permite acomodar valores novos e

emergentes.

Algumas estratégias de redução de ruído como utilizar a visão de fora ou agregar julgamentos tem a função de reduzir o ruído sem impedir a evolução dos valores. As próprias regras rígidas, podem ser programadas para empreender mudanças com o tempo.

Muitas vezes regras claras induzem a um comportamento oportunista ou interesseiro. Como regras possuem fronteiras claras, as pessoas podem evadir-se a elas procedendo a uma conduta que é tecnicamente isenta, mas acarreta danos iguais ou análogos.

Quando o objetivo é desestimular uma conduta imprópria, a imprevisibilidade ajuda mesmo que o ruído seja grande. Para saber se um sistema ruidoso impõe maior ou menor dissuasão, precisamos saber se os reais infratores são avessos ou afeitos risco. Se queremos aumentar a dissuasão, é melhor aumentar a punição mas eliminar o ruído.

A tentativa de reduzir o ruído pode sufocar motivação, afetar a criatividade, engajamento e impedir que as pessoas façam grandes descobertas. Gente criativa precisa de espaço, é necessário haver uma margem para manobra. Ao mesmo tempo que pessoas que estão em posição de julgar não gostam que sua discricionariedade lhes seja tirada, podem se sentir diminuídas e até humilhadas.

## REGRAS OU PADRÕES

Para reduzir o ruído ou decidir como e em que medida o fazer , precisamos diferenciar dois modos de regular o comportamento: regras e padrões. Regras são feitas para eliminar a discricionariedade por parte de quem as aplica, já os padrões são feitos para conceder tal discricionariedade. As regras estabelecem que não se pode exceder tal limite de velocidade, os padrões determinam que as pessoas devem dirigir “*com prudência*”.

A intenção da aplicação das regras é reduzir o ruído, elas reduzem o papel do julgamento. Os julgadores tem menos trabalho a fazer, seguem as regras e sua margem de manobra fica menor. Quando a regra é medida por um número (*limite de velocidade menor que 100 km/h*) a interpretação se torna mais fácil e o ruído é reduzido.

Padrões vão além da apuração de fatos (*regras*), é preciso atribuir conteúdo a definições vagas, um juiz talvez tenha que elaborar inúmeros julgamentos antes de decidir o que pode ser considerado “*razoável*” ou “*viável*”. Quando estabelecemos padrões , exportamos a autoridade da tomada de decisão para os outros, delegamos poder.

Diretrizes podem ser regras e padrões, quando são regras reduzem drasticamente o julgamento, quando

são padrões não passam a ser abrangentes mas não proíbem o exercício da discricionariedade. Algoritmos por exemplo, funcionam como regras.

Organizações de todo o tipo com frequência escolhem entre uma das duas opções ou a combinação de ambos. Sempre que esses órgãos estão divididos com relação a algum assunto, será mais fácil produzir padrões do que regras. Estabelecer padrões sem especificar os detalhes pode gerar ruído. Como já visto anteriormente, esse ruído pode ser controlado por meio de algumas estratégias de higiene de ruído como: julgamentos agregados, protocolo de avaliações mediadoras.

Mesmo gerando algum ruído, as organizações tem mais facilidade de criar padrões que regras pela falta de informação que as capacitaria para produzir regras sensatas. Nesse caso o melhor a fazer é criar algum tipo de padrão e confiar nos especialistas para especificar seu significado mesmo que a consequência seja um ruído.

Uma empresa como o Facebook pode dizer aos seus funcionários que derrubem o conteúdo quando o post viola uma regra clara (*nudez, por exemplo*), ou bullying (*padrão*). Receberam inúmeras queixas dizendo que os padrões geravam ruído excessivo trazendo injustiças e erros. Ao implementar os Padrões da Comunidade do Facebook em 2020, a empresa procurou definir temas como o discurso de ódio que

geraram muito ruído. A fim de diminuir, foi criado um documento com regras chamado Padrões de Implementação, nele continham fotos sobre os assuntos e a orientação clara do que fazer ao ter contato com tais temas.

Em muitas organizações o ruído é produto da criação de regras falhas. Quando o ruído é suficientemente alto, todos percebem que pessoas em situação similar não estão tendo tratamento similar. Esse movimento pode se transformar em um clamor público, uma auditoria de ruído normalmente o precede.

Justiça burocrática é o exercício de eliminar julgamentos ruidosos como a Matriz de Invalidez (*conjunto de julgamentos mecânicos baseados em: formação escolar localização geográfica e capacidades físicas remanescentes*) essa matriz promete diminuir as decisões ruidosas e injustas na área dos auxílios financeiros do governo.

Sempre que uma instituição adota decisões delimitadas por regras, corre risco de ver o ruído voltar a emergir. As regras podem produzir terríveis resultados em casos particulares, quando isso acontece o juiz pode simplesmente ignorá-las, exercendo seu poder discricionário mediante uma forma branda de desobediência civil. Tentar controlar o ruído através de regras rígidas é um alerta para a possibilidade de que as regras simplesmente mandem o poder discricionário para a clandestinidade. Quando essas

coisas acontecem, é difícil de regulamentar e até mesmo enxergar. Para evitar isso, as regras devem ser monitoradas e se houver ruído, pode ser uma pista de que elas devem ser revistas.

A escolha entre regras e padrões quando não é intuitiva, pode ser definida a partir de dois parâmetros: o custo das decisões e o custo dos erros. Com os padrões, o custo das decisões pode ser elevado já que cabe aos juízes trabalharem para lhes atribuir conteúdo. Quando médicos lidam com diretrizes claras para decidir se um paciente tem faringite, por exemplo, suas decisões podem ser rápidas e relativamente inequívocas. Com regras o custo das decisões são muito mais rápidos. Antes de aplicar uma regra, alguém tem que decidir em que ela consiste portanto criar uma regra pode ser difícil.

O custo dos erros se refere à quantidade e magnitude dos equívocos. Se os subalternos são informados e praticam higiene de decisão, um padrão pode servir perfeitamente, mesmo havendo um pouco de ruído. Se os superiores desconfiam de seus subalternos, não forem competentes, ou parciais, não podendo viabilizar a higiene de decisão, devem ser restringido por regras.

Sempre que numerosas decisões precisam ser tomadas, pode haver muito ruído, sendo melhor implementar regras, assim como em decisões repetidas.

## REVISÃO E CONCLUSÃO

Ruído é a variabilidade indesejada de julgamentos. O julgamento é uma forma de medição em que o instrumento é a mente humana, integra informalmente diversas informações separadas afim de avaliar, atribuir uma pontuação a um objeto. Certos julgamentos são preditivos, alguns verificáveis (*quando no final podemos verificar o resultado*) outros inverificáveis como previsões a longo prazo e questões fictícias. Alguns julgamentos são avaliativos (*não podem ser comparados a um valor real, objetivo*).

Classificamos como viés, quando a maioria dos erros em um conjunto de julgamentos vai na mesma direção. Ruído de sistema é o ruído observado nas organizações que empregam profissionais intercambiáveis na tomada de decisão, como médicos em pronto-socorro, juízes proferindo sentenças criminais e corretores numa companhia de seguros.

Erro quadrático médio (*EQM*) é o padrão de precisão em medições científicas, ele produz a média amostral como uma estimativa desviesada da média populacional. Trata erros positivos e negativos igualmente e penaliza de forma desproporcional erros grandes. É apropriado para fazer julgamentos preditivos em que a meta seja a precisão objetiva. Ele não é capaz de refletir os reais custos dos erros de julgamento, que são erros assimétricos.

O viés medido em um conjunto de julgamentos verificáveis é a diferença entre o julgamento médio de um caso e o valor real correspondente. Essa comparação é impossível para julgamentos não verificáveis.

O ruído em um sistema pode ser avaliado em uma auditoria de ruído (*experimento em que diversos profissionais elaboram julgamentos sobre o mesmo caso*). As auditorias podem medir a variabilidade dos julgamentos nos sistemas, as vezes elas chamam atenção para as deficiências de habilidade e treinamento. Ser ativamente receptivo (*possuir habilidades específicas na tarefa, inteligência e certo estilo cognitivo*) faz com que algumas pessoas sejam melhores julgadores.

Em algumas situações a variabilidade entre correntes é uma coisa positiva, produzem ideias, inovação mas em questões de julgamento, o ruído de sistema é sempre um problema.

O ruído de nível é a variabilidade dos julgamentos médios feitos por diferentes indivíduos. Já o ruído de padrão é a diferença nas respostas pessoais, idiossincráticas, dos juízes para um mesmo caso. O ruído de padrão estável reflete a singularidade dos juízes: suas respostas aos casos são tão individuais quanto suas personalidades. O ruído de padrão nem sempre é estável, também possui um componente transitório, esse é o caso do ruído de ocasião. Esse tipo

de ruído ocorre quando um mesmo juiz julga o mesmo caso de maneira diferente devido a influencia da ocasião como por exemplo um juiz se torna mais leniente na segunda-feira depois que o seu time venceu o jogo no domingo.

A ignorância objetiva desempenha um grande papel nos erros de julgamento, ela se caracteriza pela confiança exagerada que os juízes tem ao realizar julgamentos preditivos subestimando sua ignorância objetiva (*o fato de realmente não possuírem informações consistentes*) e seus vieses.

A estratégia de Desenviesamento propõe detectar os vieses em tempo real , designando um observador de decisão para identificar seus indícios.

A higiene de decisão é a principal sugestão do livro para eliminar ruídos, seu objetivo é prevenir os erros mesmo sem saber quais são. Os seis princípios abaixo definem esse procedimento:

O objetivo do julgamento é a precisão, não a expressão individual.

Pense estatisticamente e adote a visão de fora.

Estruture os julgamentos em diversas tarefas independentes.

Resista a intuições prematuras.

Obtenha julgamentos independentes de múltiplos juízes, depois considere agregá-los.

Prefira julgamento relativos e escalas relativas.