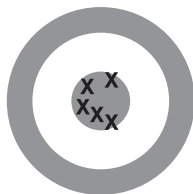


Wprowadzenie

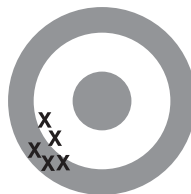
Dwa rodzaje błędów

Wyobraźmy sobie, że cztery grupy znajomych poszły na strzelnicę. Każdy z zespołów składa się z pięciu osób i strzela z tego samego karabinu, a każda osoba oddaje jeden strzał. Rysunek 1 pokazuje ich trafienia.

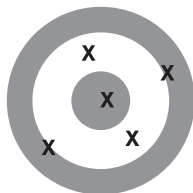
W idealnej sytuacji każdy strzał trafiłby w środek tarczy.



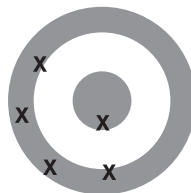
ZESPÓŁ A



ZESPÓŁ B



ZESPÓŁ C



ZESPÓŁ D

Rysunek 1. Cztery zespoły

Zespołowi A prawie się udało. Trafienia tego zespołu skupiają się ciasno wokół środka tarczy, czyli są bliskie ideału.

O strzałach zespołu B powiemy, że cechuje je tendencyjność, bo chybiają celu w sposób systematyczny. Jak widać na rysunku, odchylenie jest spójne, dzięki czemu daje się przewidywać. Gdyby ktoś z tego zespołu miał strzelać jeszcze raz, można by się było założyć, że trafienie wypadnie w tym samym obszarze co pięć poprzednich. Spójność tej tendencji zarazem każe się doszukiwać wyjaśnienia przyczynowo-skutkowego: może karabin tego zespołu ma skrzywiony celownik?

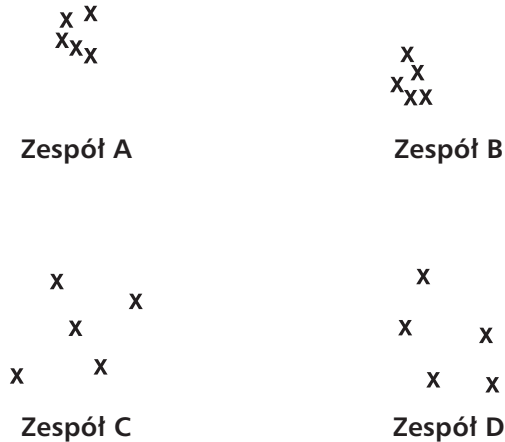
O strzałach zespołu C powiemy, że są z a s z u m i o n e (czyli dotknięte szumem), bo trafienia są rozrzucone po całej tarczy. Nie widać tu żadnej wyraźnej tendencyjności, bo trafienia są z grubsza wypośrodkowane na tarczy. Gdyby któryś z członków zespołu oddał kolejny strzał, nie wiedzielibyśmy za bardzo, gdzie może trafić. Co więcej, nie przychodzi nam do głowy żadna interesująca hipoteza na wytłumaczenie wyników zespołu C. Wiemy, że jego członkowie to kiepscy strzelcy. Nie wiemy, dlaczego w ich wynikach jest tyle szumu.

W wynikach zespołu D widać zarówno tendencyjność, jak i szum. Podobnie jak w zespole B, strzały chybiają celu w sposób systematyczny; podobnie jak w zespole C, trafienia są szeroko rozrzucone po tarczy.

Jednak ta książka nie jest o strzelectwie. Jej tematem są ludzkie błędy. Tendencyjność oraz szum — czyli błędy systematyczne oraz losowy rozrzut — są różnymi składnikami błędu. Nasze tarcze strzelnicze ilustrują tę różnicę.

Strzelnica to metafora możliwych błędów w ludzkich osądach, zwłaszcza w najróżniejszych decyzjach i osądach formułowanych w imieniu organizacji. W takich sytuacjach spotykamy oba rodzaje błędów przedstawione na rysunku 1. Niektóre osądy cechuje tendencyjność: chybiają celu w sposób systematyczny. Inne osądy są zaszumione, bo ludzie, którzy powinni się ze sobą zgadzać, ostatecznie trafiają w zupełnie różne miejsca na tarczy. Niestety, wiele organizacji boryka się zarówno z tendencyjnością, jak i z szumem.

Rysunek 2 ilustruje ważną różnicę pomiędzy tendencyjnością a szumem. Widać na nim, jak wyglądałyby tarcze zespołów, gdyby widoczna była tylko ich odwrotna strona, bez zaznaczenia środka celu.



Rysunek 2. Widok odwrotnej strony tarczy

Patrząc na odwrotną stronę tarczy, nie da się powiedzieć, który zespół — A czy B — trafił celniej. Za to na pierwszy rzut oka widać, że trafienia zespołów C i D są zaszumione, a zespołów A i B — nie. Mało tego, rysunek 2 mówi nam tyle samo o rozrzucie trafień, ile rysunek 1. Ogólną właściwością szumu jest to, że można go rozpoznawać i mierzyć, nie wiedząc nic ani o celu, ani o tendencyjności.

Ta ogólna właściwość jest dla nas kluczowa, bo wiele wniosków w tej książce pochodzi z badań poświęconych osądom, dla których odpowiedź prawidłowa jest nieznaną albo wręcz niepoznawalną. Kiedy lekarze stawiają różne diagnozy u tego samego pacjenta, to da się badać różnicę zdań pomiędzy nimi, nawet jeśli nie wiemy, co choremu dolega naprawdę. Kiedy kierownic-

two wytwórni filmowej szacuje wielkość rynku dla danego filmu, możemy badać rozrzut ich szacunków, nie wiedząc, ile film rzeczywiście zarobił, ani nawet czy w ogóle powstał. Nie musimy wiedzieć, kto ma rację, żeby zmierzyć, w jakim stopniu różnią się od siebie osądy w danej sprawie. Żeby zmierzyć szum, wystarczy popatrzeć na odwrotną stronę tarczy.

Chcąc rozumieć błędy popełniane w osądach, trzeba zrozumieć zarówno tendencyjność, i szum. Jak się przekonamy, czasami szum jest wręcz ważniejszym problemem. Jednak w debacie publicznej na temat ludzkich błędów — a także w organizacjach jak świat długi i szeroki — szum zauważa się rzadko. Tendencyjność to gwiazda pierwszego planu, a szum gra rolę epizodyczną i najczęściej pozostaje za kulisami. Tendencyjność spowodowana błędami poznawczymi jest przedmiotem tysięcy artykułów naukowych i dziesiątek książek popularnonaukowych, ale mało które z nich wspominają o szumie. Ta książka jest próbą przywrócenia równowagi.

Poziom szumu w rzeczywistych decyzjach jest często szokująco wysoki. Oto kilka przykładów niepokojącej ilości szumu w sytuacjach, w których liczy się dokładność:

- **SZUM JEST W MEDYCYNIE.** Mając do czynienia z tym samym pacjentem, różni lekarze formułują różne osądy — czy chodzi o nowotwory skóry, nowotwory piersi, choroby serca, gruźlicę, zapalenie płuc, depresję czy mnóstwo innych schorzeń. Szum jest szczególnie wysoki w psychiatrii, gdzie oczywiście ważne są osądy subiektywne. Jednak znaczny stopień szumu znajdujemy również w tak nieoczekiwanych obszarach jak interpretacja zdjęć rentgenowskich.
- **SZUM DOTYKA ORZECZEŃ O OPIECE NAD DZIEĆMI.** Pracownicy ośrodków pomocy społecznej oceniają, czy dzieci spotyka złe traktowanie w domu rodzinnym, a jeśli tak, to czy powinny trafić pod pieczę zastępczą. Decyzje poszczególnych pracowników są zaszumione w tym sensie, że jedni pracownicy skierowaliby dziecko do rodziny zastępczej, a inni nie. Po latach więcej dzieci bezceremonialnie

posłanych do rodzin zastępczych dotyczą problemy: niższe zarobki, a także wyższy odsetek przestępczości i cięż nastoletnich.

- SZUM JEST W PROGNOZACH. Zawodowi progności potrafią się bardzo różnić w prognozach sprzedaży nowego produktu, wzrostu stopy bezrobocia, ryzyka bankructwa spółki borykającej się z problemami — właściwie wszystkiego. Progności nie tylko nie zgadzają się pomiędzy sobą, ale nawet sami ze sobą. Na przykład kiedy deweloperów oprogramowania przy dwóch różnych okazjach poproszono o oszacowanie czasu niezbędnego do ukończenia zadania, ich własne szacunki wymaganej liczby godzin różniły się średnio o 71%.
- SZUM JEST W DECYZJACH DOTYCZĄCYCH AZYLANTÓW. Decyzja o przyznaniu azylu w Stanach Zjednoczonych to coś na kształt loterii. Kiedy zbadano wnioski o azyl losowo przydzielane różnym sędziom, okazało się, że jeden sędzia przyznawał azyl w pięciu procentach przypadków, a inny — w 88%. Tytuł artykułu opisującego badanie mówi sam za siebie: badacze pisali w nim o „uchodźczej ruletce” (ruletka będzie się tu przewijać jeszcze nieraz).
- SZUM JEST W DECYZJACH O ZATRUDNIENIU. Osoby prowadzące rozmowy o pracę bardzo różnie oceniają tych samych kandydatów. Znaczne rozbieżności pojawiają się również w ocenach okresowych, które bardziej zależą od osoby dokonującej oceny niż od rzeczywistych wyników w pracy.
- SZUM JEST W DECYZJACH O ZWOLNIENIU ZA KAUCJĄ. To, czy oskarżona osoba zostanie zwolniona za kaucją, czy też zaczeka na proces w areszcie, w dużej mierze zależy od osoby sędziego. Jedni sędziowie są znacznie bardziej pobłażliwi od drugich. Sędziowie znacznie różnią się też w ocenie tego, u którego oskarżonego istnieje najwyższe ryzyko ucieczki lub recydywy.
- SZUM JEST W KRYMINALISTYCE. Nauczono nas myśleć, że identyfikacja na podstawie analizy odcisków palców jest niezawodna. Jednak czasami eksperci od daktyloskopii różnią się w ocenie, czy odcisk znaleziony na miejscu zbrodni

pasuje do linii papilarnych podejrzanego. Eksperti nie tylko nie zgadzają się ze sobą nawzajem — eksperci czasami podejmują różne decyzje, kiedy dwukrotnie oceniają te same odciski. Podobną zmienność udokumentowano też w innych obszarach kryminalistyki, łącznie z analizą DNA.

- **SZUM DOTYKA DECYZJI PATENTOWYCH.** Jak piszą autorzy najważniejszego badania poświęconego temu zagadnieniu, „o tym, czy urząd patentowy przyzna patent, czy odrzuci wnioski, w istotnym stopniu decyduje to, który pracownik ocenia wnioski”. Naturalnie rodzi to niepokojące pytania na temat sprawiedliwości takich decyzji.

Wszystko to zaledwie czubek dużej góry lodowej. Gdziekolwiek spojrzymy na ludzkie osądy, prawdopodobnie znajdziemy w nich szum. Żeby formułować lepsze osądy, trzeba przewyciężyć nie tylko błędy poznawcze i uprzedzenia wywołujące tendencyjność, ale także szum.

Nasza książka składa się z sześciu części. W części pierwszej przyjrzymy się różnicy między szumem a tendencyjnością oraz wykazemy, że szum dotyka zarówno organizacji publicznych, jak i firm prywatnych — i to nieraz w stopniu szokującym. Żeby docenić skalę problemu, na początek przyjrzymy się osądom w dwóch obszarach: jednym są wyroki wydawane w procesach karnych (sektor publiczny), drugim ubezpieczenia (sektor prywatny). Na pierwszy rzut oka te dwa obszary różni wszystko, jednak pod względem szumu łączy je wiele. Żeby to wykazać, wprowadzamy pojęcie „audytu szumu”, czyli narzędzia mającego określać rozbieżności w decyzjach fachowców rozpatrujących te same sprawy w obrębie tej samej organizacji.

W części drugiej przyglądamy się naturze ludzkich osądów oraz kwestii mierzenia ich trafności i popełnianych błędów. Osądy są podatne zarówno na tendencyjność, jak i na szum. W tej książce opisujemy uderzającą równowagę obu rodzajów błędów. Szum sytuacyjny to rozbieżności w osądach formułowanych przez jedną osobę czy grupę w tej samej sprawie, tyle że w różnych sytuacjach. Dyskusje w grupie rodzą zaskakująco

wiele szumu w wyniku z pozoru nieistotnych czynników, takich jak kolejność zabierania głosu.

W części trzeciej przyglądamy się bliżej pewnemu rodzajowi osądów, który doczekał się wielu badań: osądom prognozytycznym (przewidywaniom i prognozom). Omawiamy kluczową przewagę reguł, formuł i algorytmów nad osądami ludzkimi przy tworzeniu prognoz: wbrew powszechnemu przekonaniu jest ona kwestią nie tyle lepszego zrozumienia sprawy, co wyeliminowania szumu z osądów. Piszemy o nieprzekraczalnej granicy trafności osądów prognozytycznych, spowodowanej obiektywną niewiedzą na temat przyszłości, a także pokazujemy, jak taka niewiedza oraz szum ograniczają jakość naszych przewidywań. Na koniec mierzymy się z pytaniem, które z pewnością zdąży się nasunąć i tobie: skoro szum jest tak wszechobecny, dlaczego nie zauważyliśmy go wcześniej?

W części czwartej zwracamy się ku ludzkiej psychologii, żeby wyjaśnić zasadnicze przyczyny szumu. Należą do nich różnice między ludźmi wynikające z różnych czynników, w tym z różnic w osobowości i stylu poznawczym, różnych stopni wagi przypisywanej przez osoby poszczególnym aspektom sytuacji, a także tego, że ludzie na różne sposoby posługują się tą samą skalą. Zastanawiamy się, dlaczego pozostajemy nieświadomi szumu i dlaczego zdarzenia albo osądy często nas nie dziwią, pomimo że były całkowicie nie do przewidzenia.

W części piątej przyglądamy się kwestii praktycznej: podnoszeniu jakości własnych osądów i zapobieganiu błędom. (Czytelnicy i czytelniczki, których interesuje głównie wykorzystanie redukcji szumu w praktyce, mogą przejść od razu do tej części z pominięciem omówienia wyzwań związanych z prognozowaniem oraz kwestii psychologicznych w częściach trzeciej i czwartej). Tutaj przyglądamy się działaniom ograniczającym szum w medycynie, biznesie, edukacji, administracji państwowej i na innych polach. Przedstawiamy kilka technik ograniczania szumu, które zbiorczo nazywamy higieną decyzyjną. Prezentujemy pięć różnych studiów przypadku w dziedzinach, w których badania wykazały istnienie znacznego poziomu

szumu i gdzie podjęto konsekwentne działania dla jego ograniczenia — z różnym skutkiem, co także wiele mówi na temat problemu. Badane przypadki dotyczą: zawodnych diagnoz medycznych; ocen okresowych pracowników; kryminalistyki; decyzji o zatrudnianiu pracowników oraz prognozowania jako takiego. Na koniec tej części przedstawiamy system, który nazywamy metodą analiz pośrednich (MAP): uniwersalne podejście do oceniania opcji, wykorzystujące kilka kluczowych praktyk higieny decyzyjnej i pozwalające formułować rzetelniejsze, mniej dotknięte szumem osądy.

Jaki poziom szumu jest właściwy? Z tą kwestią mierzymy się w części szóstej. Może to sprzeczne z intuicją, ale właściwy poziom szumu nie jest zerowy. W niektórych obszarach wyeliminowanie szumu jest po prostu niewykonalne, w innych — zbyt drogie, a w jeszcze innych działania na rzecz ograniczenia szumu odbijałyby się niekorzystnie na innych ważnych wartościach: na przykład mogłyby podkopywać morale i rodzić w ludziach poczucie, że są traktowani jak trybiki w maszynie. Częściowe włączenie algorytmów do rozwiązania tego problemu również rodzi wiele zastrzeżeń; w tej części przyjrzymy się niektórym z nich. Jednak tak czy inaczej, aktualny poziom szumu jest nie do przyjęcia. Apelujemy zarówno do organizacji prywatnych, jak i publicznych o przeprowadzenie „audytów szumu” i podejścia do działań mających ograniczyć szum z bezprecedensową powagą. Pozwoliłoby to złagodzić powszechną niesprawiedliwość — a przy okazji zmniejszyć koszty w wielu obszarach.

Z myślą o tej aspiracji każdy rozdział zamykamy kilkoma krótkimi stwierdzeniami w formie cytatów. Możesz wykorzystać te sformułowania słowo w słowo albo dostosować je do dowolnych kwestii, które są dla Ciebie ważne, czy to związanych ze zdrowiem, czy z bezpieczeństwem, edukacją, pieniędzmi, pracą, rozrywką, czy jeszcze innymi sprawami. Zrozumienie problemu szumu oraz próby jego rozwiązania to zbiorowy i wciąż jeszcze niedokończony wysiłek. Wszyscy mamy szansę się do niego przyczynić. Ta książka powstała z nadzieją, że zdołamy tę szansę wykorzystać.

Część pierwsza

Jak znaleźć szum

Trudno się pogodzić z sytuacją, że podobni ludzie skazywani za takie samo przestępstwo otrzymują diametralnie różne wyroki — dajmy na to, że jedna osoba idzie na pięć lat do więzienia, a druga dostaje wyrok w zawieszeniu. A jednak podobne rzeczy dzieją się w wielu miejscach. Oczywiście wymiar sprawiedliwości karnej jest również dotknięty problemem tendencyjności w formie uprzedzeń i stronniczości. Jednak tu, w rozdziale 1, skupimy się na szumie — a w szczególności na tym, jak pewien słynny sędzia zwrócił społeczną uwagę na jego istnienie, uznał taki stan rzeczy za skandal i zainicjował krucjatę, która w pewnym sensie zmieniła świat (choć nadal w niedostatecznym stopniu). Nasza opowieść toczy się w Stanach Zjednoczonych, ale podobne historie na pewno da się opowiedzieć (i kiedyś się opowie) o wielu innych krajach. Podejrzewamy, że są kraje, w których problem jest jeszcze poważniejszy niż w Stanach Zjednoczonych. Sięgamy po ten przykład częściowo po to, żeby pokazać, jak szum może się stawać przyczyną ogromnej niesprawiedliwości.

Orzecznictwo w sprawach karnych to kwestia wyjątkowo dramatyczna, ale interesuje nas tu również sektor prywatny, gdzie stawka też bywa wysoka. Ilustrujemy to w rozdziale 2 przykładem dużego towarzystwa ubezpieczeniowego, w którym aktuariusze ustalają wysokość składek ubezpieczeniowych dla potencjalnych klientów, a likwidatorzy szkód muszą oceniać wartość roszczeń o odszkodowania. Można by sądzić, że takie zadania powinny być proste i mechaniczne, a różni pracownicy

będą podawać z grubsza jednakowe kwoty. Żeby się przekonać, czy tak jest rzeczywiście, przeprowadziliśmy starannie przygotowany eksperyment: audyt szumu. Wyniki nas zaskoczyły, ale — co ważniejsze — zdumiały i skonsternowały także kierownictwo firmy. Jak się przekonałiśmy, rozmiary szumu obecnego w decyzjach pracowników oznaczają dla towarzystwa ubezpieczeniowego duże koszty. Ten przykład pokazuje, że szum może powodować poważne straty gospodarcze.

Oba przykłady dotyczą badań prowadzonych na dużej liczbie osób formułujących dużą liczbę osądów. Jednak wiele istotnych osądów to nie są oceny powtarzalne, lecz jednorazowe: dotyczą takich kwestii, jak z pozoru wyjątkowa okazja biznesowa czy wprowadzenie na rynek nowego produktu, walka z pandemią czy zatrudnienie osoby nie mieszczącej się w standardowym profilu pracownika. Czy szum daje się znaleźć również w tak unikatowych sytuacjach? Mogłoby się wydawać, że nie — przecież szum to inaczej niepożądana rozbieżność zdań, a skąd miałyby się brać rozbieżności w wyjątkowych decyzjach? Na to pytanie próbujemy odpowiedzieć w rozdziale 3. Twój osąd — nawet formułowany w sytuacji z pozoru wyjątkowej — jest zaledwie jedną z chmury możliwości. Tutaj również pojawia się wiele szumu.

Wniosek wyłaniający się z tych trzech rozdziałów można podsumować jednym zdaniem, które będzie motywem przewodnim tej książki: „tam, gdzie jest osąd, tam też jest szum — i jest go więcej, niż myślisz”. Spróbujmy się więc przyjrzeć, ile go właściwie jest.

Zbrodnia i zasumiona kara

Założmy, że ktoś został skazany za przestępstwo — kradzież w sklepie, posiadanie heroiny, napaść albo rozbój z bronią w rękę. Jaki prawdopodobnie zapadnie wyrok?

Odpowiedź nie powinna zależeć od tego, który sędzia orzeka w sprawie, czy danego dnia jest zimno czy gorąco albo czy miejscowa drużyna sportowa wygrała wczoraj mecz. Byłoby czymś oburzającym, gdyby trzy podobne osoby otrzymały zupełnie różne kary za to samo przestępstwo: jedna wyrok w zawieszeniu, druga dwa lata pozbawienia wolności, a trzecia — dziesięć lat. A jednak taką oburzającą sytuację spotyka się w wielu krajach — nie tylko w odległej przeszłości, ale i dzisiaj.

Na całym świecie sędziom od dawna pozostawiano dużą dowolność w ustalaniu wymiaru kary. W wielu krajach eksperci chwalą taką uznaniowość jako rzecz sprawiedliwą i humanitarną. Utrzymują, że wyroki karne powinny być wymierzane w zależności od wielu czynników: nie tylko samego przestępstwa, ale również charakteru oskarżonego i okoliczności. Obowiązującym dogmatem były wówczas kary dobierane indywidualnie: jeśli zwiążemy sędziom ręce przepisami, sprawców spotykać będzie odhumanizowane traktowanie, bo przestaną być postrzegani jako jednostki mające prawo wskazać na szczegóły własnej sytuacji. W oczach wielu sama idea „uczciwego procesu” zdawała się domagać pozostawienia sędziom marginesu swobody.

W latach siedemdziesiątych XX w. powszechny entuzjazm dla sędziowskiej swobody zaczął się załamywać z prostego powodu: alarmujących danych ukazujących szum. W 1973 roku uwagę

społeczną zwrócił na ten problem słynny sędzia federalny Marvin Frankel. Przed otrzymaniem nominacji sędziowskiej Frankel był obrońcą wolności słowa i praw człowieka, współzałożycielem organizacji ochrony praw człowieka Lawyers Committee for Human Rights (obecnie znanej jako Human Rights First).

Frankel potrafił walić prawdę prosto w oczy, a szum w systemie sprawiedliwości karnej uważał za bulwersujący. Oto jak opisywał własne motywy:

Osoba oskarżona o napad na bank skazywana na podstawie przepisów prawa federalnego mogła dostać wyrok do 25 lat więzienia. Mogło to oznaczać dowolny wyrok od 0 do 25 lat. A jak się szybko przekonałem, ostateczny wymiar kary zależał nie tyle od danej sprawy czy osoby oskarżonego, ile od konkretnego sędziego, czyli jego poglądów, predylekcji i uprzedzeń. Zatem ta sama osoba oskarżona w tej samej sprawie mogła otrzymać bardzo różne kary w zależności od tego, któremu sędziemu przydzielono jej proces.

Frankel nie przytoczył na poparcie tych stwierdzeń żadnych statystycznych analiz, za to pokazał wiele barwnych przykładów anegdotycznych, ilustrujących nieuzasadnione rozbieżności w traktowaniu podobnych osób. Dwaj mężczyźni, obaj niekarani, zostali skazani za wprowadzenie w obieg sfałszowanych czeków — jeden na kwotę 58,40 dolarów, drugi 35,20 dolarów. Pierwszego skazano na 15 lat, drugiego na 30 dni pozbawienia wolności. W podobnych procesach o defraudację jeden mężczyzna został skazany na 117 dni więzienia, drugi na 20 lat. Frankel wskazał na wiele podobnych przypadków, ubolewając nad — jak to określił — „rozległymi i niemal całkowicie niekontrolowanymi uprawnieniami” sędziów federalnych, skutkującymi „arbitralnymi okrucieństwami popełnianymi każdego dnia”, co uznawał za rzecz nie do przyjęcia w kraju podlegającym „rządom nie ludzi, ale prawa”.

Frankel wezwał amerykański Kongres do położenia kresu „dyskryminacji”, jak określił przejawy arbitralnego okrucieństwa. Mówiąc to, miał na myśli przede wszystkim szum

w postaci niewytłumaczalnych rozbieżności w wyrokach. Jednak przedmiotem jego troski była również tendencyjność, to znaczy uprzedzenia w formie nierówności rasowych i socjoekonomicznych. Jak twierdził, żeby walczyć zarówno z szumem, jak i tendencyjnością, potrzebny był zakaz nierównego traktowania oskarżonych, chyba że różnice dałoby się „uzasadnić odpowiednimi testami, które dałoby się formułować i stosować z obiektywnością dającą pewność, że wyniki nie będą kwestią uznaniowych ukazów ze strony urzędników, sędziów czy innych osób” (przez „uznaniowy ukaz” Frankel rozumiał uzależnienie decyzji od osobistego kaprysu). Co więcej, Frankel domagał się ograniczenia szumu za pomocą „szczegółowych profili lub *checklist* czynników, w miarę możliwości opartych na ocenach numerycznych lub innych obiektywnych metodach”.

Frankel, który pisał to na początku lat siedemdziesiątych, nie posuwał się wprawdzie do promowania idei „zastąpienia ludzi maszynami” (jak to sam określił), ale był tego niepokojąco bliski. Jego zdaniem „praworządność wymaga zbioru bezosobowych i powszechnie obowiązujących zasad, które odnosiłyby się nie tylko do sędziów, ale do wszystkich”. Wyraźnie opowiadał się za wykorzystaniem „komputerów jako narzędzia, które pomogłoby uporządkować myślenie o orzecznictwie w prawie karnym”. Rekomendował też utworzenie specjalnej komisji ds. orzecznictwa.

Książka Frankela stała się jedną z najbardziej wpływowych pozycji w historii prawa karnego — nie tylko w Stanach Zjednoczonych, ale na całym świecie. Owszem, nie była doskonała, bo brakowało jej formalnej dyscypliny. Była druzgocąca w wymowie, ale nie miała oparcia w naukowym rygorze. Aby się przekonać, czy rzeczywiście szum jest realnym problemem, kilkoro badaczy natychmiast podjęło jego idee i zaczęło analizować poziom szumu w wyrokach karnych.

Jedno z pierwszych zakrojonych na szeroką skalę badań tego rodzaju, któremu przewodniczył sam Frankel, przeprowadzono w 1974 roku. Pięćdziesięcioro sędziów z różnych rejonów kraju miało w hipotetycznych scenariuszach wydawać wyroki

w sprawie tych samych oskarżonych i na podstawie identycznych opisów. Głównym ustaleniem płynącym z badania było to, że „normą był brak konsensusu”, a różnice pomiędzy orzekanymi karami były „porażające”. W zależności od osoby sędziego ten sam dealer heroiny mógł dostać rok albo dziesięć lat. Kara za napad na bank mogła wynieść pięć lat więzienia, ale mogła i osiemnaście. W sprawie o wymuszenie w badaniu padł zarówno drastyczny wyrok dwudziestu lat pozbawienia wolności i 65 tysięcy dolarów grzywny, jak i zaledwie trzech lat więzienia bez kary grzywny. Co najbardziej alarmujące, w szesnastu z dwudziestu spraw sędziowie nie zgadzali się nawet co do tego, czy w ogóle należy wymierzyć karę pozbawienia wolności.

Po tym badaniu przyszła seria kolejnych, ale we wszystkich stwierdzano podobnie szokujący poziom szumu. Na przykład w 1977 roku William Austin i Thomas Williams przeprowadzili badanie kwestionariuszowe w grupie 47 sędziów, którzy mieli zarekomendować wyroki w pięciu jednakowych sprawach dotyczących drobnych przestępstw. Dla każdej ze spraw przedstawiono im streszczenia prawdziwych ustaleń procesowych, na których opierali się sędziowie wydający rzeczywiste wyroki, w tym informacje o zarzutach, treści zeznań, ewentualnej karalności i statusie społecznym oskarżonych, a także o ich osobistym charakterze. Kluczowe ustalenie badania wskazywało na „znaczące rozbieżności”. Na przykład w sprawie kradzieży z włamaniem zalecane wyroki pozbawienia wolności mieściły się w zakresie od pięciu lat do zaledwie 30 dni (plus grzywna 100 dolarów). W sprawie posiadania marihuany jedni sędziowie rekomendowali pozbawienie wolności, inni wyrok w zawieszeniu.

W znacznie większym badaniu z 1981 roku przyjrzano się grupie 208 sędziów federalnych, którzy mieli rozpatrywać szesnaście jednakowych hipotetycznych spraw. Najważniejsze ustalenia były porażające:

W zaledwie trzech z szesnastu spraw sędziowie byli jednomyślni co do zasadności kary pozbawienia wolności. Nawet gdy większość zgadzała się co do zastosowania kary więzienia, między sędziami

nadal istniały duże różnice pod względem długości zalecanych kar. W procesie o oszustwo, gdzie średni wymiar orzekanych kar wynosił 8,5 roku więzienia, najdłuższym z zaleconych wyroków było dożywocie. W innej sprawie średni wymiar kary wynosił 1,1 roku więzienia, ale najdłuższym wyrokiem było 15 lat.

Choć wyniki te były bardzo wymowne, a badania opierały się na starannie kontrolowanych eksperymentach, to i tak niemal na pewno zaniżają one rzeczywiste rozmiary szumu w wymiarze sprawiedliwości. Prawdziwi sędziowie mają do czynienia ze znacznie większą liczbą informacji niż uczestnicy badań, którzy dostawali tylko starannie dobrane szkice na temat podsądnych. Część takich dodatkowych informacji jest oczywiście istotna, ale wiele wskazuje na to, że również informacje nieistotne — w postaci drobnych, z pozoru losowych czynników — mogą się przekładać na duże różnice w rozstrzygnięciu sprawy. Stwierdzono na przykład, że sędziowie są bardziej skłonni udzielić więźniowi przedterminowego zwolnienia na początku dnia pracy albo po przerwie na posiłek niż tuż przed taką przerwą. Kiedy sędziowie są głodni, robią się surowsi.

Badanie, które objęło tysiące wyroków w sądach dla nieletnich, wykazało, że gdy lokalna drużyna futbolowa ponosi w weekend porażkę, w poniedziałek sędziowie będą orzekać surowiej (zjawisko utrzymuje się również przez pozostałe dni tygodnia, choć staje się słabsze). Brzemień tej zwiększonej surowości sędziów spada w przeważającej mierze na czarnych oskarżonych. W innym badaniu przyjrano się zbiorowi półtora miliona decyzji sądowych z trzech dekad — tu też stwierdzono, że w dniach po porażce lokalnego zespołu futbolowego sędziowie orzekają surowiej niż w dniach po zwycięstwie.

Badanie obejmujące sześć milionów orzeczeń wydanych przez sędziów we Francji w okresie dwunastu lat wykazało, że oskarżonych traktuje się łagodniej w urodziny (mowa o urodzinach oskarżonych — podejrzewamy, że sędziowie mogą być pobłażliwsi również we własne urodziny, jednak o ile wiemy, tej hipotezy nie badano). Na sędziów mogą wpływać nawet

czynniki tak nieistotne, jak temperatura na zewnątrz. Przegląd 207 000 decyzji sądów imigracyjnych wykazał znaczące efekty tak nieistotnych czynników jak codzienne wahania temperatury: kiedy jest upał, ludzie rzadziej otrzymują azyl. Jeśli w ojczystym kraju spotykają cię polityczne prześladowania i szukasz azylu za granicą, trzymaj kciuki, a może wręcz módl się, żeby twoje posiedzenie przypadło w chłodny dzień.

Jak zmniejszyć szum w wyrokach

W latach siedemdziesiątych argumenty Frankela i potwierdzające je ustalenia empiryczne trafiły na biurko Edwarda M. Kennedy'ego, brata zabitego w zamachu prezydenta Johna F. Kennedy'ego i jednego z najbardziej wpływowych amerykańskich senatorów. Kennedy był wstrząśnięty i zbulwersowany. Już w 1975 roku wprowadził pierwszą inicjatywę legislacyjną zmierzającą do reformy orzecznictwa w prawie karnym. Reforma spełzała na niczym, ale Kennedy nie odpuszczał. Powołując się na dane, rok po roku naciskał na uchwalenie nowych przepisów. W 1984 roku odniósł sukces: w obliczu danych o bezzasadnych różnicach w wydawanych wyrokach amerykański Kongres uchwalił ustawę o reformie orzecznictwa.

Nowe przepisy miały ograniczyć szum w systemie, redukując „nieskrępowaną uznaniowość, którą prawo pozostawia sędziom i członkom komisji ds. zwolnień przedterminowych odpowiadającym za wydawanie i wdrażanie wyroków”. W szczególności Kongres odniósł się do „bezzasadnie szerokich” rozbieżności w wydawanych wyrokach, powołując się konkretnie na wyniki badań mówiące, że w rejonie Nowego Jorku rzeczywisty wymiar kar orzekanych w identycznych sprawach potrafił się wahać od trzech do dwudziestu lat więzienia. Ustawa przewidywała utworzenie Amerykańskiej Komisji ds. Orzecznictwa, której domagał się Frankel, a która otrzymała jasne zadanie: wydawać wiążące wytyczne dla wyroków mające zmniejszyć dopuszczalny zakres kar.

W następnym roku komisja wydała wytyczne, zasadniczo oparte na średniej kar wymierzanych za podobne przestępstwa, wyliczonej na podstawie analizy dziesięciu tysięcy autentycznych spraw. Mocno zaangażowany w ten proces sędzia Sądu Najwyższego USA, Stephen Breyer, bronił decyzji o oparciu się na wcześniejszych wyrokach, wskazując na nierozwiązywalny spór w obrębie samej komisji: „Dlaczego Komisja, zamiast oprzeć się na historii, nie spróbowała wypracować własnych decyzji na podstawie racjonalnych przesłanek? Krótko mówiąc: bo nie byliśmy w stanie tego zrobić. Nie byliśmy w stanie, bo w dowolnej sprawie dałoby się wysunąć dobre, acz wzajemnie sprzeczne argumenty. [...] Spróbujcie uszeregować listę wszystkich możliwych przestępstw pod względem wymiaru kary, na jaką zasługują. [...] Potem zbierzcie takie listy od znajomych i przekonajcie się, czy będą takie same. Od razu mogę powiedzieć, że nie będą”.

Wytyczne nakazują sędziom uwzględniać w wyrokach dwa czynniki: samo przestępstwo oraz kryminalną przeszłość oskarżonego. W zależności od powagi czynu, przestępstwu przypisuje się jeden z czterdziestu trzech „poziomów”. Kryminalna przeszłość oskarżonego dotyczy głównie liczby oraz wagi wcześniejszych skazań. Po uwzględnieniu poziomu przestępstwa i kryminalnej przeszłości oskarżonego wytyczne pozostawiają stosunkowo wąskie widełki możliwych wyroków: najwyższy dopuszczalny wymiar kary może przekraczać najniższy o maksymalnie sześć miesięcy albo 25% (w zależności od tego, która liczba okaże się większa). Sędzia może wymierzyć karę spoza widełek, jeśli widzi szczególne okoliczności łagodzące albo obciążające, ale takie odstępstwa należy uzasadnić przed sądem apelacyjnym.

Choć wytyczne są obowiązkowe, to jednak nie są sztywne. Nadal daleko im do tego, czego pragnął Frankel, bo pozostawiają sędziom znaczące pole do manewru. Jednak z wielu badań wykorzystujących różne metody i skupiających się na różnych okresach płynie ten sam wniosek: wytyczne zmniejszają szum. Albo, mówiąc bardziej technicznym językiem, „zredukowały

poziom zróżnicowania w orzekanych karach dający się przypisać czynnikowi losowemu, jakim jest tożsamość orzekającego sędziego”.

Najbardziej rozbudowane badanie przeprowadziła sama komisja. Porównano w nim wyroki wydawane w sprawach napadów na bank, handlu kokainą, handlu heroiną oraz malwersacji bankowych, zestawiając z jednej strony wyroki wydawane w 1985 roku (przed wejściem w życie wytycznych), a z drugiej — wyroki z okresu od 19 stycznia 1989 roku do 30 września 1990 roku. Sprawców porównywano, uwzględniając wszystkie czynniki, które w wytycznych uznano za istotne przy wymierzeniu odpowiedniego wymiaru kary. Dla każdego przestępstwa zróżnicowanie pomiędzy karami wymierzonymi przez poszczególnych sędziów było znacznie mniejsze w późniejszym z okresów, czyli po wdrożeniu wytycznych.

Inne badanie wykazało, że spodziewana różnica pomiędzy karami wymierzonymi przez różnych sędziów wynosiła 17% (4,9 miesiąca) w latach 1986-1987. W latach 1988-1993 różnica ta spadła do 11% (3,9 miesiąca). Niezależne badanie obejmujące inne okresy czasowe wykazało podobną skuteczność w ograniczaniu rozbieżności pomiędzy sędziami, definiowanych jako różnica pomiędzy średnim wymiarem kar wymierzonych przez sędziów orzekających w podobnej liczbie spraw.

Pomimo tych ustaleń wytyczne spotkały się z żywiołową krytyką. Niektórzy (w tym wielu sędziów) byli zdania, że część wyroków jest zbyt surowa — jednak taka uwaga dotyczy tendencji, a nie szumu. Z naszego punktu widzenia w tej książce znacznie ciekawsze zastrzeżenie wysunęło wielu sędziów, którzy twierdzili, że wytyczne są głęboko niesprawiedliwe, bo nie pozwalają sędziom należycie uwzględnić wyjątkowości konkretnej sprawy. Ceną za ograniczenie szumu okazała się większa mechaniczność podejmowanych decyzji, co uznawano za rzecz nie do przyjęcia. Jak napisali prof. Kate Stith (Yale) i sędzia federalny José Cabranes, „sprawiedliwość nie wymaga ślepoty, tylko wnikliwego spojrzenia i równego traktowania”, co „może mieć

miejsce jedynie wtedy, kiedy orzeczenie uwzględni złożoność konkretnej sprawy”.

Za tym zarzutem poszły energiczne wysiłki w celu podważenia wytycznych, czy to na gruncie prawnym, czy politycznym. Początkowo wysiłki te spełzły na niczym, jednak w 2005 roku Sąd Najwyższy USA uchylił wytyczne z powodów zupełnie niezwiązanych z debatą, o której tutaj piszemy. To orzeczenie sprawiło, że wytyczne nabrały charakteru konsultacyjnego, czyli stały się czysto orientacyjne. Zwłaszcza sędziowie federalni przyjęli orzeczenie z dużym zadowoleniem. 75% sędziów wołało wytyczne konsultacyjne, a tylko 3% uważało wytyczne obowiązkowe za lepsze.

Jaki był skutek zmiany charakteru wytycznych z obowiązkowego na konsultacyjny? Tę kwestię zbadała profesor Crystal Yang z wydziału prawa Uniwersytetu Harvarda, choć nie za pomocą eksperymentu czy ankiety, a ogromnego zbioru danych w postaci autentycznych wyroków w sprawach dotyczących niemal 400 tysięcy oskarżonych. Jej główne ustalenie pokazuje, że po 2005 roku rozbieżności pomiędzy sędziami — mierzone na wiele sposobów — znacząco wzrosły. Kiedy wytyczne były obowiązkowe, kary zasądzone przez relatywnie surowych sędziów były o 2,8 miesiąca dłuższe, niż gdy orzekali sędziowie średnio surowi. Kiedy wytyczne stały się jedynie orientacyjne, ta rozbieżność wzrosła dwukrotnie. Komentarz Yang do złudzenia przypomina słowa sędziego Frankela sprzed czterdziestu lat: „te ustalenia budzą poważne wątpliwości co do sprawiedliwości wyroków, skoro osoba orzekającego sędziego w znacznym stopniu przyczynia się do różnicy w traktowaniu podobnych oskarżonych skazywanych za podobne przestępstwa”.

Po tym, jak wytyczne nabrały charakteru konsultacyjnego, sędziowie częściej kierowali się osobistymi wartościami — obowiązkowe wytyczne ograniczają nie tylko szum, ale i tendencyjność. Po decyzji Sądu Najwyższego znacząco wzrosła rozbieżność pomiędzy wyrokami wymierzonymi za to samo przestępstwo Afroamerykanom i białym. Zarazem sędzie

częściej niż sędziowie wykorzystywały zwiększoną swobodę wyboru do wydawania łagodniejszych wyroków. To samo można powiedzieć o sędziach nominowanych przez prezydentów z Partii Demokratycznej.

Uchylenie obowiązkowych wytycznych trzy lata po śmierci sędziego Frankela w 2002 roku przywróciło stan bardziej przypominający jego koszmar: prawo bez porządku.

* * *

W historii walki sędziego Frankela o wprowadzenie wytycznych w orzecznictwie przewija się kilka głównych kwestii, o których będziemy pisać w tej książce. Po pierwsze, formułowanie osądów jest trudne, bo rzeczywistość jest skomplikowana i niepewna. Jest to oczywiste w wymiarze sprawiedliwości, ale też w większości sytuacji wymagających osądów w formie „fachowej oceny”, czyli — ogólnie rzecz biorąc — decyzji podejmowanych w medycynie, pielęgniarstwie, systemie prawnym, inżynierii, szkolnictwie, architekturze, przemyśle filmowym, komisjach rekrutacyjnych, wydawnictwach, wszelkiego rodzaju kierownictwach biznesowych, prowadzeniu klubów sportowych czy w wielu innych dziedzinach. Tam, gdzie pojawia się osobisty osąd, rozbieżność zdań staje się nieunikniona.

Po drugie, rozbieżność zdań jest znacznie większa, niż się spodziewamy. Co do zasady, mało kto jest przeciwny pozostawieniu sędziom możliwości dokonania „roztropnej oceny”, za to prawie każdy źle odbiera skalę rozbieżności, jakie się pojawiają w wyniku takich ocen. *S u m s y s t e m o w y*, czyli niepożądane rozbieżności w osądach, które w idealnej sytuacji powinny być identyczne, może powodować rażącą niesprawiedliwość, wysokie koszty oraz najróżniejsze błędy.

Po trzecie, szum można redukować. Podejście zalecane przez sędziego Frankela, które wdrożyła amerykańska komisja ds. orzecznictwa — oparte na regułach i wytycznych — to tylko jeden z kilku sposobów skutecznego ograniczania szumu. Przy innych rodzajach osądów lepiej sprawdzają się inne

podejścia. Niektóre podejścia ograniczające szum redukują zarazem tendencyjność.

Po czwarte, próby zmniejszenia szumu często budzą protesty i napotykać poważne trudności. Bez rozwiązania również tych kwestii walka z szumem kończy się porażką.

Jak mówić o szumie w orzecznictwie karnym

„Eksperymenty pokazują duże rozbieżności pomiędzy wyrokami zalecanymi przez różnych sędziów w identycznych sprawach. To nie może być sprawiedliwe. Wyrok nie powinien zależeć od tego, na którego sędziego trafi oskarżony”.

„Wyrok w sprawie karnej nie powinien zależeć od humoru sędziego w danym dniu albo od temperatury na zewnątrz”.

„Jednym ze sposobów rozwiązania tej sprawy są wytyczne, ale wielu osobom wytyczne się nie podobają, bo ograniczają swobodę decyzji sędziów, a ta może być konieczna, jeśli wyroki mają być sprawiedliwe i trafne. W końcu każda sprawa jest wyjątkowa, prawda?”