

GÜRÜLTÜ

İnsan Yargısında Hata

DANIEL KAHNEMAN

OLIVIER SIBONY

CASS R. SUNSTEIN

PEGASUS

"Üretken olmayan politikaların en büyük kaynağı sanılanın aksine ön yargılar, yolsuzluk ve kötü niyet değil, 'sezgi, cehalet ve atalet'tir. Bu üç kötülüğün ne kadar yaygın olduğunu ve onlarla savaşmak için neler yapabileceğimizi çok net bir şekilde ortaya koyan *Gürültü*, çok temel ve ufuk açıcı bir okuma sunuyor."

Prof. Dr. Esther Duflo,
2019 Nobel Ekonomi Ödüllü Yazar

"Yazarlar, insan yargısındaki kusurlara dair benzersiz ve yeni içgörülerini insan çabasının her alanına zekice uyguluyorlar... *Gürültü*, psikoloji alanında bir dönüm noktası yaratıyor."

Prof. Dr. Philip E. Tetlock

"İnsanları nasıl değerlendirdiğinizi, nasıl karar verdiğinizi ve sorunları nasıl çözeceğinizi yeniden düşünmenize yardımcı olacak dünyanın en büyük beyinlerinden üçünü okumaya hazır olun."

Prof. Dr. Adam Grant

"*Gürültü*, gözlerden uzak bir yerde saklanan devasa bir toplumsal soruna yönelik mükemmel bir incelemedir."

Prof. Dr. Steven Levitt

"Kahneman, Sibony ve Sunstein bir fil kadar büyük bir problem keşfettiler: gürültü. Bu ilham verici kitapta bize gürültünün neden önemli olduğunu, neden fark ettiğimizden çok daha fazla olduğunu ve onu nasıl azaltabileceğimizi gösteriyorlar. Onların tavsiyelerini uygulamak bize daha kârlı işler, daha sağlıklı vatandaşlar, daha adil bir hukuk sistemi ve daha mutlu hayatlar sağlayacaktır."

Prof. Dr. Jonathan Haidt

Pegasus Yayınları: 2606

Gürültü
Daniel Kahneman
Olivier Sibony
Cass R. Sunstein
Özgün Adı: Noise: A Flaw in Human Judgment

Editör: Onur Gayretli
Düzelti: Mustafa Erdoğan
Sayfa Tasarımı: Ezgi Gültekin

Baskı-Cilt: Alioğlu Matbaacılık
Sertifika No: 45121
Orta Mah. Fatin Rüştü Sok. No: 1/3-A
Bayrampaşa/İstanbul
Tel: 0212 612 95 59

1. Baskı: İstanbul, Eylül 2023
ISBN: 978-625-410-389-6

Türkçe Yayın Hakları © PEGASUS YAYINLARI, 2023
Copyright © Daniel Kahneman, Olivier Sibony, Cass R. Sunstein, 2021

Tüm hakları saklıdır. Bu kitapta yer alan fotoğraf/resim ve metinler Pegasus Yayıncılık Tic. San. Ltd. Şti.'den izin alınmadan fotokopi dâhil, optik, elektronik ya da mekanik herhangi bir yolla kopyalanamaz, çoğaltılamaz, basılamaz, yayımlanamaz.

Yayıncı Sertifika No: 45118

Pegasus Yayıncılık Tic. San. Ltd. Şti.
Gümüşsuyu Mah. Osmanlı Sk. Alara Han
No: 11/9 Taksim / İSTANBUL
Tel: 0212 244 23 50 (pbx) Faks: 0212 244 23 46
www.pegasusyayinlari.com / info@pegasusyayinlari.com

pegasusyayinlari pegasusyayinevi pegasusyayinlari Pegasus Yayınları

DANIEL KAHNEMAN
OLIVIER SIBONY
CASS R. SUNSTEIN

GÜRÜLTÜ

İnsan Yargısında Hata

İngilizceden çeviren:
Sevinç Seyla Tezcan

PEGASUS YAYINLARI

Noga, Ori ve Gili için... DK
Fantin ve Lélia için... OS
Samantha için... CRS

İçindekiler

Giriş: İki Tür Hata 9

KISIM 1: Gürültüyü Bulmak 19
1. Suç ve Gürültülü Ceza 21
2. Gürültülü Bir Sistem 32
3. Tekil Kararlar 45

KISIM 2: Zihniniz Bir Ölçü Aletidir 51
4. Yargı Meseleleri 55
5. Hatayı Ölçmek 68
6. Gürültü Analizi 83
7. Durum Gürültüsü 94
8. Gruplar Gürültüyü Nasıl Artırır? 111

KISIM 3: Öngörücü Yargılarda Gürültü 127
9. Yargılar ve Modeller 131
10. Gürültüsüz Kurallar 144
11. Objektif Bilgisizlik 159
12. Normal Vadisi 170

KISIM 4: Gürültü Nasıl Olur? 183
13. Kısayollar, Ön Yargılar ve Gürültü 185
14. Eşleştirme Operasyonu 201
15. Ölçekler 213
16. Örüntüler 227
17. Gürültü Kaynakları 238

KISIM 5: Yargıları İyileştirmek 251
18. Daha İyi Yargılar için Daha İyi Yargıçlar 255
19. Ön Yargıları Gidermek ve Karar Hijyeni 268
20. Adli Tıpta Bilgi Dizilimi 278
21. Tahminde Seçme ve Bir Araya Toplama 294
22. Tıpta Kılavuzlar 310
23. Performans Puanlamalarında Ölçeği Tanımlamak 326
24. İşe Alımda Yapılandırma 341
25. Aracı Değerlendirmeler Protokolü 355

KISIM 6: Optimal Gürültü 371
26. Gürültüyü Azaltmanın Maliyetleri 377
27. Saygınlık 388
28. Kurallar mı Standartlar mı? 400

Gözden Geçirme ve Sonuç: Gürültüyü Ciddiye Almak 413
Son Söz: Daha Az Gürültülü Bir Dünya 431

EK A: Gürültü Denetimi Nasıl Yapılır? 433
EK B: Karar Gözlemcisi Bir Kontrol Listesi 441
EK C: Tahminleri Düzeltmek 445
Teşekkürler 451
Notlar 453
Dizin 501

GİRİŞ

İki Tür Hata

Dört arkadaş takımın bir poligona gittiğini düşünün. Her takım beş kişiden oluşuyor, aynı tüfeği paylaşıyorlar ve herkes bir atış yapıyor. Sonuçlar Şekil 1'deki gibidir.

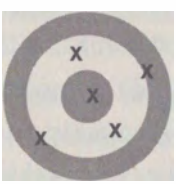
İdeal bir dünyada, her atış hedefi tam on ikiden vururdu.



TAKIM A



TAKIM B



TAKIM C



TAKIM D

ŞEKİL 1: *Dört takım*

Takım A için durum aşağı yukarı budur. Takımın atışları, kusursuza yakın bir örüntüyle hedef tahtasının ortasında toplanıyor.

Takım B için *ön yargılı* diyoruz çünkü atışları sistematik olarak hedefi şaşırmış. Şekilde görüldüğü gibi ön yargının tutarlılığı, bir tahmini destekliyor: Şayet takımda biri bir atış daha yapacak olsaydı, o atışın ilk beş atışla aynı alana isabet edeceği üzerine bahse girebilirdik. Ön yargının tutarlılığı, ayrıca nedensel bir açıklamaya da davetiye çıkarıyor: Muhtemelen takımın tüfeğinin nişangâhı eğriydi.

Takım C için *gürültülü* diyoruz çünkü atışları geniş bir alana yayılmış. Bariz bir ön yargı yok çünkü atışlar hedef tahtası üstünde kabaca ortada toplanmış. Takımdan biri bir atış daha yapacak olsaydı nereye isabet edebileceği konusunda elimizde çok çok az bilgi olurdu. Dahası akla Takım C'nin sonuçlarını açıklayacak ilginç bir hipotez gelmiyor. Takım üyelerinin kötü nişancılar olduğunu biliyoruz. Neden bu kadar gürültülü olduklarını ise bilmiyoruz.

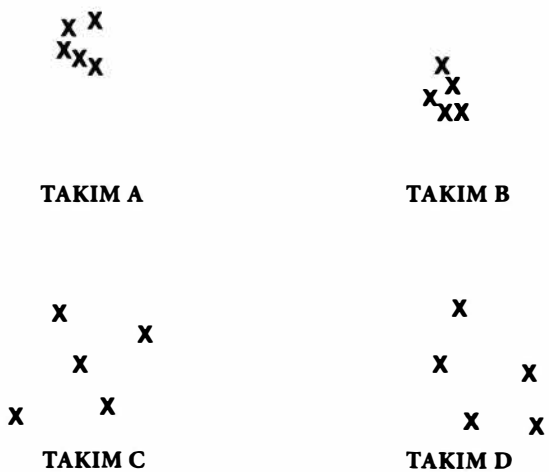
Takım D ise hem ön yargılı hem gürültülüdür. Takım D'nin atışları da Takımı B gibi sistematik olarak hedefi şaşırmış ve Takım C gibi geniş bir alana yayılmış.

Takdir edersiniz ki bu kitabın konusu hedef tahtasına nasıl atış yapılması gerektiği değildir. Konumuz insan hataları. Ön yargı ve gürültü –sistematik sapma ve rastgele saçılım– hatanın farklı ögeleridir. Hedefler, farkı görselleştiriyor.

Poligon, insan yargısında, özellikle insanların organizasyonlar adına aldıkları çeşitli kararlarda çıkabilecek sorunlar için bir mecaz görevi üstleniyor. Bu örneklerde, Şekil 1'de resmedilen iki hata türünü bulacağız. Bazı yargılar, ön yargılıdır; sistematik olarak hedefi şaşırırlar. Diğer yargılar ise hemfikir olması beklenen insanların kendilerini hedefin etrafında çok farklı noktalarda bulması nedeniyle gürültülüdür. Ne yazık ki pek çok organizasyon hem ön yargı hem gürültüden muztariptir.

Şekil 2, ön yargı ve gürültü arasındaki önemli bir farkı göstermektedir. Poligonda, elinizde nişan aldıklarına dair herhangi bir gösterge yokken, size sadece takımların ateş ettikleri hedefin arka yüzü gösterilse, ne göreceğinizi ortaya koyuyor.

Hedef tahtasının arkasına baktığınızda Takım A'nın mı yoksa Takım B'nin mi hedefi tam on ikiden vurmaya daha yakın olduğunu ayırt edemezsiniz. Ama Takım C ve D'nin gürültülü olduğunu, Takım A ve B'nin öyle olmadığını bir bakışta anlayabilirsiniz. Hatta saçılma hakkında Şekil 1'deki kadar bilgi sahibi olursunuz. Gürültünün genel bir özelliği, onu, hedef ya da ön yargı hakkında hiçbir şey bilmeden tanıyabilmeniz ve ölçebilmenizdir.



ŞEKİL 2: *Hedef tahtasının arkasına bakış*

Gürültünün az önce bahsettiğimiz genel özelliği, buradaki amaçlarımız açısından oldukça önemli zira vardığımız sonuçların birçoğu, gerçek cevabı bilinmeyen ve hatta bilinemez olan yargılardan çekildi. Hekimler aynı hastaya farklı teşhis koyduklarında, hekimler arasındaki bu görüş ayrılıklarını, hastaya neyin rahatsızlık verdiğini bilmeden inceleyebiliriz. Film yapımcıları bir

film için piyasa tahmini yaptığında, cevaplarının değişkenliğini filmin, sonunda ne kadar gişe yaptığını ve hatta çekilip çekilmediğini bile bilmeden inceleyebiliriz. Aynı davanın yargılarının ne kadar değişkenlik sergilediğini ölçmek için kimin haklı olduğunu bilmemiz gerekmez. Gürültüyü ölçmek için tek yapmamız gereken, hedef tahtasının arkasına bakmaktır.

Yargıda hatayı anlamak için hem ön yargıyı hem gürültüyü anlamalıyız. Bazen, göreceğimiz gibi, gürültü daha önemli bir sorundur. Ancak insan hatası hakkında halk arasında ve dünyanın dört bir yanındaki organizasyonlarda yapılan konuşmalarda, gürültü nadiren dikkate alınır. Ön yargı şovun yıldızıdır. Gürültü genellikle sahne dışında kalan figürandır. Ön yargı konusu binlerce bilimsel makalede ve düzinelerce popüler kitapta ele alınmasına rağmen çok azı gürültü meselesine değinir. Bu kitap, aradaki dengeyi yeniden sağlama girişimimiz olacak.

Gerçek dünyada alınan kararlarda gürültü miktarı, genellikle skandal düzeyde yüksektir. Hassasiyetin ve doğruluğun önemli olduğu durumlarda gürültünün telaş verici miktarına örnek verecek olursak:

Tıp gürültülüdür. Farklı doktorlar aynı hastayla karşı karşıya kaldıklarında hastanın cilt kanseri, meme kanseri, kalp hastalığı, tüberküloz, zatürre, depresyon ve daha birçok hastalığa sahip olup olmadığı konusunda farklı yargılara varırlar. Gürültü, tahmin edileceği gibi öznel yargının önemli olduğu psikiyatride özellikle yüksektir. Ancak, röntgenlerin yorumlanması gibi hiç beklenmeyen alanlarda da hatırı sayılır miktarda gürültüye rastlanabilir.

Çocuk velayeti kararları gürültülüdür. Çocuk koruma dairelerindeki vaka yöneticileri, çocukların istismar edilme riski altında olup olmadığını, ortada risk varsa çocukların koruyucu aileye verilip verilmemesi gerektiğini değerlendirmekten so-

rumludur. Yöneticilerin bir kısmının çocukları koruyucu aile yanına vermeye diğerlerinden daha meyilli olması nedeniyle, sistem gürültülüdür. Seneler sonra, bu zalim yöneticiler tarafından koruyucu aile yanına gönderilen talihsiz çocukların büyük bir oranı –daha yüksek suç işleme oranları, daha yüksek ergenlik çağı doğum oranları, daha düşük kazanç gibi– daha olumsuz şartlarla karşımıza çıkar.

Öngörüler gürültülüdür. Profesyonel öngörücüler, bir ürünün olası satış miktarından işsizlik oranında olası artışa ve sorunlu şirketler için iflas olasılığına kadar hemen her konuda hayli değişken öngörülerde bulunurlar. Sadece birbirleriyle değil, kendileriyle de fikir ayrılığına düşerler. Örneğin aynı yazılımcılara aynı görevin tamamlanma süresi için tahminleri iki ayrı günde sorulduğunda, öngördükleri süreler ortalama %71 oranında değişiklik göstermiştir.

Sığınma kararları gürültülüdür. Sığınma talebinde bulunan bir kişinin ABD'ye kabul edilip edilmeyeceği bir tür piyangoya bağlıdır. Rastgele yargıçlara dağıtılan vakalar incelendiğinde, bir yargıç başvuruların sadece %5'ini onaylarken, bir başkasının %88'ini onayladığı görülmüştür. Konuyla ilgil bir çalışmanın başlığı her şeyi anlatıyor: "Sığınmacı Ruleti." (Bol miktarda rulet göreceğiz.)

Personel kararları gürültülüdür. İş başvuru mülakatlarını yürüten kişiler aynı insanlar üstünde çok farklı değerlendirmeler yaparlar. Bazı çalışanların performans puanlamaları da çok değişkendir ve değerlendirilen performanstan çok değerlendirmeyi yapan kişiye dayanır.

Kefalet kararları gürültülüdür. Suçlanan kişiye kefalet hakkı tanınacak mı, yoksa kişi duruşmayı beklerken hapse gönderilecek mi sorusunun cevabı, kısmen davayı dinleyen yargıcın kimliğine bağlıdır. Kimi yargıçlar diğerlerinden çok daha

merhametlidir. Yargıçların hangi davalılarda kaçma ya da yeniden suç işleme riskinin daha yüksek olduğu konusundaki değerlendirmeleri de dikkat çekici ölçüde farklılık arz eder.

Adli tıp gürültülüdür. Bize parmak iziyle teşhisin şaşmaz olduğu öğretildi. Ancak parmak izi incelemesi yapan uzmanlar, bazen bir suç mahallinde bulunan bir izin zanlınınkine uyup uymadığına karar verirken farklı yargılarda bulunabiliyorlar. Uzmanlar sadece birbirleriyle görüş ayrılığı yaşamıyor; bazen aynı uzman ona aynı parmak izinin sunulduğu farklı durumlarda tutarsız kararlar verebiliyor. Benzer değişkenlik adli tıbbın başka disiplinlerinde, hatta DNA analizinde bile belgelenmiştir.

Patent verme kararları gürültülüdür. Patent başvuruları üzerine yapılmış öncü bir çalışmanın yazarları, işin içinde gürültü olduğunu vurguluyorlar: "Patent ofisinin bir patenti verip vermeyeceği, şans eseri uygulamaya hangi müfettişin verildiğine göre değişir." Bu değişkenlik, tahmin edileceği gibi, hakkaniyet açısından sorun yaratmaktadır.

Bütün bu gürültülü durumlar, kocaman bir buz dağının sadece görünen kısmıdır. İnsan yargılarına baktığınız her yerde gürültü bulma ihtimaliniz çok yüksektir. Yargılarımızın kalitesini artırmak için ön yargı kadar gürültüyü de yenmeliyiz.

Elinizdeki çalışma altı kısımdan oluşuyor. Kısım 1'de gürültü ile ön yargı arasındaki farkı ele alacak ve gerek kamusal gerekse özel organizasyonların zaman zaman şoke edici boyutlarda gürültülü olabileceğini göstereceğiz. Problemin ne olduğunu kavramak için, iki alandaki yargılarla başlayacağız. Birincisi cezai hükümleri (ve dolayısıyla kamu sektörünü) içeriyor. İkincisi ise sigortacılığı (dolayısıyla özel sektörü) ele alıyor. Bu iki alan, ilk bakışta birbirinden ancak bu kadar farklı görünebilir. Öte yandan

gürültü açısından ciddi anlamda ortak noktaları bulunuyor. Bu noktanın doğruluğunu kanıtlamak için, bir organizyon içinde yer alan benzer vakaları ele alırken, profesyoneller arasında ne kadar çok görüş ayrılığı yaşandığını ölçmek için tasarlanmış gürültü denetimi fikrini öne süreceğiz.

Kısım 2'de insan yargısının doğasını derinlemesine ele alarak doğruluğu ve hatayı nasıl ölçeceğimizi keşfedeceğiz. Yargılar, ön yargıya olduğu kadar gürültüye de açıktır. İki hata türünün rollerinde çarpıcı bir denklik tanımlıyoruz. Durumsal gürültü aynı vaka konusunda, aynı insan ya da grup tarafından farklı durumlarda varılan yargılardaki değişkenliktir. Önce kimin söz aldığı gibi görüntüde alakasız faktörler yüzünden grup tartışmalarından şaşırtıcı miktarlarda durum gürültüsü doğar.

Kısım 3'te kapsamlı olarak araştırılan bir yargı türünü daha derinlemesine ele alacağız: Öngörücü yargı. Öngörülerde bulunmak konusunda kuralların, formüllerin ve algoritmaların insanlar karşısındaki kilit avantajının, popüler inancın aksine kuralların üstün içgörüsünden çok gürültüleri olduğunu göreceğiz. Öngörücü yargının kalitesinin nihai sınırını –geleceğe dair nesnel bilgisizlik– ve öngörünün kalitesini sınırlamak için gürültüyle nasıl iş birliği yaptığını anlatacağız. Son olarak, muhtemelen o noktada kendinize çoktan sormuş olacağınız bir soruya odaklanacağız: Gürültü bu kadar sık rastlanan bir şeyken, onu daha önce nasıl fark etmediniz?

Kısım 4'te insan psikolojisine yönelecek ve gürültünün merkezî nedenlerini açıklayacağız. Bu nedenler, kişilik ve bilişsel tarzın, farklı değerlendirmelerin tartılmasında kişiye özgü farklılıkların ve insanların aynı ölçeklerden farklı şekillerde faydalanmasının da aralarında olduğu çeşitli faktörlerden kaynaklanan kişiler arası farklılıkları içeriyor. İnsanların gürültüye neden kayıtsız kaldığını ve öngörmelerinin mümkün olmadığı olay ve yargılara sıklıkla neden şaşırmadıklarını masaya yatıracağız.

Kısım 5'te, yargılarınızı nasıl iyileştirebileceğiniz ve hatayı nasıl önleyebileceğiniz sorusunu ele alıyoruz. (Öncelikli olarak pratik gürültü azaltma uygulamalarıyla ilgilenen okuyucular Kısım 3 ve 4'teki öngörü zorlukları ve yargı psikolojisi tartışmalarını geçip bu kısımdan başlayabilirler.) Tıp, iş dünyası, eğitim, hükûmet gibi kurumlarda gürültüyü alt etme çabalarını soruşturacağız. Sizi *karar hijyeni* etiketi altında topladığımız gürültü azaltma teknikleriyle tanıştıracağız. Gürültünün belgelendiği ve insanların gürültüyü azaltmak için sürekli bir çaba sergilediği, farklı başarı seviyeleriyle sonuçlanan beş vaka incelemesi sunacağız. Bu vaka incelemeleri arasında tutarsız tıbbi teşhisler, performans puanlamaları, adli tıp, işe alım kararları ve genel olarak öngörüler yer alıyor. Bu kısmı *aracı değerlendirmeler protokolü* adını verdiğimiz bir sistemle, seçeneklerin değerlendirilmesine, karar hijyeninin çeşitli kilit pratiklerini bünyesinde barındıran ve daha az gürültülü, daha güvenilir yargılar üretmeyi hedef alan genel amaçlı bir yaklaşımla tamamlayacağız.

Doğru gürültü seviyesi nedir? Kısım 6'da bu konuyu ele alıyoruz. Belki de genel kanının aksine doğru seviye sıfır değildir. Bazı alanlarda gürültüyü ortadan kaldırmak mümkün değildir. Bazı alanlarda ise bunun ciddi bir bedeli vardır. Yine bazı alanlarda gürültüyü azaltma çabaları, önemli kimi değerlerden ödün vermeye neden olur. Örneğin gürültüyü ortadan kaldırma çabaları, moral gücünü baltalayabilir ve insanlara bir makinenin dişlileri gibi muamele gördükleri hissini verebilir. Algoritmalar cevabın bir parçası olduklarında, çok çeşitli itirazlarda bulunurlar; bu itirazların bir kısmını burada ele alacağız. Bununla birlikte, mevcut gürültü seviyesi kabul edilemez. Gerek özel gerekse kamusal organizasyonları, gürültü denetimleri yürütmeye ve gürültüyü azaltmak için benzeri görülmemiş bir ciddiyetle daha güçlü çabalar harcamaya davet ediyoruz. Bunu yapmaları hâlinde organizasyonlar yaygın haksızlıkları ve pek çok alanda maliyetleri azaltabilirler.

Bu amacı aklımızda tutarak her kısmı alıntılar şeklinde birkaç kısa öneriyle bitireceğiz. Bu beyanatları oldukları gibi kullanabilir ya da sağlık, güvenlik, eğitim, para, istihdam, eğlence gibi sizin için önemli olan meselelere uyarlayabilirsiniz. Gürültü sorununu anlamak, onu çözmeye çalışmak, daha yapılacak çok şeyin olduğu bir çalışmanın yanı sıra kolektif bir girişimdir. Hepimizin bu çabaya katkı sağlama fırsatı var. Bu kitap o fırsatları değerlendirebileceğimiz umuduyla yazıldı.

KISIM 1

Gürültüyü Bulmak

Aynı suçtan hüküm giyen benzer insanların çarpıcı derecede farklı cezalar –örneğin birinin beş yıl hapis, diğerinin şartlı tahliye– alması kabul edilemez. Bununla birlikte pek çok yerde bu tür şeyler olur. Hiç şüphesiz ceza muhakeme sistemi ön yargının da istilası altındadır. Ancak Bölüm 1'de üzerinde durduğumuz konu gürültü ve özellikle, ünlü bir yargıcın gürültüye dikkat çektiği, bunun çok büyük bir skandal olduğunu belirttiği ve dünyayı bir anlamda (ama yeterince değil) değiştirecek bir mücadele başlattığı zaman yaşananlar olacak. Hikâyemiz Amerika Birleşik Devletleri'nde geçiyor ama benzer hikâyelerin pek çok farklı ülke için de anlatılabileceğinden (ve anlatılacağından) şüphemiz yok. Bu ülkelerin bazılarında gürültü sorunu, ABD'dekinden çok daha kötü boyutta olabilir. Hüküm verme örneğini kısmen gürültünün büyük haksızlıklara yol açabileceğini göstermek için kullanıyoruz.

Ceza hükmü, özellikle büyük bir dram içerir ama biz, kayıpların çok büyük olabildiği özel sektör ile ilgileniyoruz. Bu noktayı tasvir edebilmek için Bölüm 2'de büyük bir sigorta şirketini ele alacağız. Orada risk değerlendirme uzmanları, potansiyel müşteriler için sigorta primlerini ayarlama görevini üstlenir ve sigorta eksperine de, hasarın değerini hükme bağlamak düşer. Bu görevlerin basit

ve mekanik olacağını ve farklı profesyonellerin kabaca aynı miktarlara ulaşacağını düşünebilirsiniz. Bu öngörüyü sınamak için, dikkatle tasarlanmış bir deney –bir gürültü denetimi– gerçekleştirdik. Sonuçlar bizi şaşırttı ama daha önemlisi şirket yönetimini hayrete düşürdü ve mutsuz etti. Öğrendiğimize göre gürültünün hacmi, şirkete çok büyük para kaybına neden oluyordu. Bu örneği gürültünün çok büyük ekonomik kayıplara yol açabileceğini göstermek için kullanıyoruz.

Her iki örnek de yargıya varan çok sayıda insanın incelenmesini içeriyor. Ancak bu yargıların ekseriyeti yinelenen değil benzersiz bir iş fırsatını ele almak, yeni bir ürünü piyasaya sürüp sürmemek, bir salgınla baş etmek, standart profile uymayan birini işe alıp almamak gibi *tekil* yargılardır. Bu tür benzersiz durumlarla ilgili kararlarda da gürültü bulunabilir mi? Burada gürültünün olmayacağını düşünmek cazip gelebilir. Sonuçta gürültü istenmeyen bir değişkendir öyle değil mi: Tekil kararlarda değişkenlik nasıl olabilir ki? Bölüm 3'te bu soruya cevap bulmaya çalışacağız. Vardığınız yargı, benzersiz görünen bir durumda bile bir olasılıklar bulutunun içinden bir yargıdır. Orada da çok gürültü bulabilirsiniz.

Bu üç bölümden çıkacak ana fikir, bu kitabın kilit teması olacak tek bir cümleyle özetlenebilir: *Yargının olduğu yerde gürültü vardır; üstelik tahmin ettiğinizden daha fazla.* Ne kadar olduğunu öğrenmek için hadi başlayalım.

BÖLÜM 1

Suç ve Gürültülü Ceza

Birinin –mağaza hırsızlığı, eroin bulundurmak, saldırı ya da silahlı soygun gibi– bir suçtan hüküm giydiğini düşünün. Ceza muhtemelen nasıl olur?

Bu soruya verilecek cevap, davaya atanan herhangi bir yargıca, dışarıda havanın sıcak ya da soğuk olmasına veya yerel bir futbol takımının bir önceki gün galip gelip gelmediğine bağlı olmamalıdır. Aynı suçtan hüküm giymiş üç insanın, –biri şartlı tahliye, diğeri iki yıl hapis, üçüncüsü on yıl hapis gibi– çarpıcı derecede farklı cezalar alması çok çirkin olurdu. Bununla birlikte bu çirkinlik, pek çok ülkede uzak bir geçmişte kalmadığı gibi, bugün de görülebiliyor.

Dünyanın her yerinde yargıçlar çok uzun yıllardan beri uygun cezalara karar vermek konusunda sağduyulu davranmaktalar. Pek çok ülkede uzmanlar bu sağduyuyu hem adil hem insani bularak el üstünde tutarlar. Söz konusu uzmanlar ceza hükümlerinin sadece suça değil, sanığın karakterini ve şartlarını da içeren bir dizi faktöre dayandırılması gerektiğinde ısrar ettiler. Ceza hükmünü bireysel olarak uyarlamak yaygınlık kazandı. Yargıçlar yalnızca kurallara göre hüküm verirse suçlular insanlık dışı muameleye

maruz kalırlardı; durumlarının detaylarına dikkat çekme hakkına sahip benzersiz, tekil bireyler olarak görülmezlerdi. Hukukun usulüne uygun işleyiş fikri, pek çoklarına açık uçlu takdir hakkı gerektiriyor gibi görünüyordu.

1970'lerde takdir hakkına duyulan evrensel heyecan tek bir basit nedenle, şaşırtıcı gürültü kanıtlarıyla çökmeye başladı. 1973 yılında Marvin Frankel adında ünlü bir yargıç, kamuoyunun dikkatini bu soruna çekti. Frankel yargıç olmadan önce Hukukçuların İnsanlar Hakları Komitesi'nin (Önce İnsan Hakları olarak bilinen bir örgüt) kurulmasına yardım eden, tutkulu bir ifade özgürlüğü ve insan hakları savunucusuydu.

Frankel coşkulu olabiliyordu. Dahası ceza muhakemesi sistemindeki gürültü konusunda çileden çıkmıştı. Motivasyonunu şu şekilde anlatıyordu:

> *Bir federal banka soyguncusu sanık hüküm giydiğinde, azami 25 yıl ceza alabilirdi. Bu 0 ile 25 yıl arasında herhangi bir süre anlamına gelirdi. Çok geçmeden fark ettim ki bu sürenin hangi sayıda sabitleneceği, bireysel olarak sanıktan çok yargıca, onun görüşlerine, eğilimlerine ve ön yargılarına bağlı oluyordu. Böylece aynı davada aynı sanık davaya hangi yargıcın baktığına bağlı olarak, çok farklı cezalar alabiliyordu.*

Frankel bu argümanı destekleyecek istatistiksel analizler sunmadı ama benzer insanların gördükleri muamelelerdeki gayrimeşru farklılıkları gösteren bir dizi güçlü anekdot paylaştı. Sabıka kaydı bulunmayan iki adamdan biri 58,40 dolar diğeri 35,20 dolar tutarında sahte çek bozdurma suçundan hüküm giymişti. İlk adam on beş *yıl* cezaya çarptırılırken, ikinci adam sadece 30 *gü*n ceza almıştı. Buna benzer zimmetine para geçirme davalarında, bir adam 117 *gün* hapis cezası alırken, diğeri 20 *yılla* cezalandırılmıştı. Fran-

kel buna benzer sayısız davaya dikkat çekerek federal yargıçların "neredeyse tamamen kontrolsüz ve geniş yetkileri"nin "her gün işlenen keyfî zulümler"e yol açtığını ve bunun "insanlara değil, hukuka dayalı bir hükûmet" için kabul edilemez olduğunu belirtti.

Frankel, söz konusu keyfî zulümleri tanımlarken Kongre'yi bu "ayrımcılığa" son vermeye çağırdı. Bu terimle ağırlıklı olarak kastettiği, cezalarda açıklanamaz farklılıklar şekline bürünmüş gürültüydü. Ayrıca ırksal ve sosyoekonomik eşitsizlikler şekline giren ön yargıdan da endişeliydi. Hem gürültü hem ön yargıyla mücadele için sanıklara yönelik muamelede farklılıklara müsamaha gösterilmemesi, daha açık bir ifadeyle bu farklılıklar "sonuçların resmî görevlilerin ve yargıçların nevi şahsına münhasır fermanlarından ibaret olmamasını garanti altına alacak bir nesnellikle ifade edilebilen ve uygulanabilen ilişkili testlerle meşrulaştırılmadığı sürece" izin verilmemesi konusunda ısrar etti. (*Nevi şahsına münhasır fermanlar* terimi biraz ezoterik kalıyor; Frankel bununla kişisel buyrukları kastediyordu.) Daha önemlisi Frankel, gürültünün "detaylı bir profil ya da mümkün olan her durumda, bir tür sayısal veya benzer nesnel puanlama içerecek bir faktör kontrol listesi" aracılığıyla azaltılması gerektiğini savundu.

1970'ler gibi erken bir dönemde yazdığı için olsa gerek, "insanların yerini makinelerin alması" olarak adlandırdığı şeyi savunacak kadar ileri gitmedi. Ama şaşırtıcı bir şekilde buna yaklaştı. "Hukukun üstünlüğünün, geniş kapsamlı olarak uygulanabilecek ve diğer herkes kadar yargıçları da bağlayacak bir kişiler üstü kurallar bütünü gerektirdiği"ne inanıyordu. "Bilgisayarların hüküm vermede düzen içinde düşünmeye yardımcı olarak" kullanılabileceğini açıkça ileri sürmüştü. Ayrıca hüküm verme konusunda bir komisyon kurulmasını da tavsiye etmişti.

Frankel'in kitabı, sadece Amerika'da değil tüm dünyada, ceza hukuku tarihinin en etkili kitaplarından biri oldu. Çalışmasında bir derece resmiyet eksikliği sıkıntısı vardı. Yıkıcıydı ama izle-

nime dayanıyordu. Çok sayıda insan, gürültünün gerçekliğini test etmek için vakit kaybetmeden cezai hükümde gürültü seviyesini keşfe çıktı.

Bu türden ilk geniş ölçekli çalışmaların Yargıç Frankel'in bizzat başkanlık ettiği bir örneği 1974'te yapıldı. Farklı bölgelerden elli yargıca benzer ceza öncesi raporlarla özetlenen farazi davaların sanıkları için hüküm vermeleri istendi. Temel bulgu "uzlaşmanın olmamasının kural olduğu" şeklindeydi ve cezalar arası farklılıklar "şaşırtıcı"ydı. Bir eroin satıcısı, yargıca bağlı olarak bir ile on yıl arası hapis cezası alabiliyordu. Bir banka soyguncusunun cezası, beş ile on sekiz yıl hapis arasında değişebiliyordu. Çalışma, bir şantaj vakasında cezaların yirmi sene hapis ve altmış beş bin dolar para cezası ve sadece üç yıl hapis ve sıfır para cezası arasında değiştiğini ortaya koydu. En şaşırtıcı olanı, yirmi vakanın on altısında hapis cezasının uygun olup olmadığı konusunda görüş birliği yoktu.

Bu çalışmayı, her biri şaşırtıcı gürültü seviyeleri saptayan bir dizi çalışma daha izledi. Örneğin 1977'de William Austin ve Thomas Williams kırk yedi yargıçla bir çalışma yürüterek yargıçlardan her biri düşük seviyeli suçlarla alakalı beş vaka için karar vermelerini istediler. Vakalara dair tüm tanımlar, yargıçların gerçek hükümlerde kullandığı suç, tanıklık, sanığın (varsa) sabıka kaydı, sosyal çevresi, geçmişi ve kişiliğiyle bağlantılı kanıtlar gibi bilgi özetlerini içeriyordu. Kilit bulgu, "elle tutulur miktarda ihtilaf" oldu. Örneğin soygun içeren bir vakada tavsiye edilen cezalar, beş yıl ile (yüz dolar para cezasıyla birlikte) otuz gün hapis cezası arasında değişiyordu. Marihuana bulundurmakla ilgili bir vakada, yargıçların bir kısmı hapis süreleri önerirken, diğerleri şartlı tahliye kararı vermişti.

1981 yılında yapılan çok daha geniş ölçekli bir çalışma, aynı on altı farazi vakanın önlerine konduğu iki yüz sekiz federal yargıçla yapıldı. Çalışmanın merkezî bulguları şaşırtıcıydı:

On altı vakadan sadece üçünde hapis cezası vermek konusunda oy birliği sağlandı. Çoğu yargıcın bir hapis cezasının uygun düşeceği konusunda hemfikir olduğu durumlarda bile, tavsiye edilen hapis süreleri arasında çarpıcı farklılıklar vardı. Ortalama hapis süresinin 8,5 yıl olduğu bir dolandırıcılık davasında en uzun ceza süresi müebbet hapisti. Başka bir vakada ortalama hapis süresi 1,1 yılken, tavsiye edilen en uzun hapis süresi 15 yıl oldu.

Çok sıkı kontrol edilen deneyler içeren bu çalışmalar, ziyadesiyle aydınlatıcı olmakla birlikte gerçek ceza hukuku dünyasında var olan gürültünün boyutunu aslında olduğundan daha hafif gösteriyordu. Gerçek hayatta yargıçlar, çalışmaya katılanların bu deneylerin özenle düzenlenmiş özetlerinde aldıklarından çok daha fazla bilgi sahibi olurlar. Bu ek bilgilerin bazıları elbette konuyla ilgilidir ama küçük ve görüntüde rastgele faktörler kisvesine bürünen bol miktarda alakasız bilgi, sonuçlarda önemli farklar yaratabilir. Misal yargıçların günün başlangıcında ya da yemek molasından hemen sonra şartlı tahliye kararı verme olasılığı, bu tür bir molanın hemen öncesine göre daha yüksek. Yargıçlar açken daha sert oluyorlar.

Binlerce çocuk mahkemesi kararı üstünde yapılan bir çalışma, hafta sonu yerel futbol takımının yenilmesi hâlinde, yargıçların pazartesi günü (ve gittikçe azalmakla birlikte o haftanın tamamında) daha sert kararlar aldığını ortaya koydu. Bu artan sertliğin orantısız bir şekilde en şiddetli kısmına siyah sanıklar maruz kalıyordu. Otuz yıl boyunca alınan 1,5 milyon yargı kararını inceleyen başka bir çalışmada benzer bir şekilde yargıçların yerel futbol takımının mağlubiyetini izleyen günlerde, galibiyeti izleyen günlere göre çok daha katı olduklarını ortaya koydu.

Fransa'da on iki yıl boyunca alınan altı milyon kararın incelenmesi sonucu, sanıklara doğum günlerinde daha fazla müsamaha

gösterildiği anlaşıldı. (Burada sanığın doğum gününden bahsediyoruz; yargıçların kendi doğum günlerinde daha da müsamahalı olacağını tahmin ediyoruz ama bildiğimiz kadarıyla bu hipotez henüz sınanmadı.) Dışarıdaki hava sıcaklığı gibi alakasız şeyler bile yargıçları etkileyebiliyor. Dört yıl boyunca alınan göçmenlikle alakalı iki yüz yedi bin mahkeme kararının incelenmesi, günlük hava sıcaklığının farklı kararlar verilmesi üstünde hatırı sayılır etkisi olduğunu ortaya koydu: Dışarıda hava sıcaksa, insanların sığınma hakkı alma ihtimali azalıyordu. Ülkenizde politik zulme maruz kalıyorsanız ve başka bir yerde sığınma hakkı talep etmek istiyorsanız duruşmanızın havanın serin olacağı bir güne denk gelmesini umabilir, hatta bunun için dua edebilirsiniz.

Hüküm Vermede Gürültüyü Azaltmak

1970'lerde Frankel'in argümanları ve onları destekleyen ampirik bulgular, ABD Senatosu'nun en nüfuzlu üyelerinden, suikasta kurban giden Başkan John F. Kennedy'nin kardeşi Edward M. Kennedy'nin dikkatini çekti. Gördükleri karşısında dehşete düşen Kennedy, korkmuştu. 1975 gibi erken bir tarihte ceza reformu mevzuatını gündeme getirdi ama bir yere ulaşamadı. Gelgelelim Kennedy'nin durmaya niyeti yoktu. Senelerce, kanıtları işaret ederek, o mevzuatın yürürlüğe girmesi için baskı uygulamaya devam etti. 1984 yılında başarıya ulaştı. Kongre gayrimeşru farklılıklarla ilgili kanıtlara tepki olarak 1984 Ceza Reformu Yasası'nı yürürlüğe koydu.

Yeni kanunun, "hukukun cezai hükümleri dayatmak ve uygulamakla sorumlu olan yargıç ve şartlı tahliye memurlarına bahşettiği sınırsız takdir yetkisi"ni azaltarak sistemdeki gürültüyü zayıflatması amaçlanmıştı. Kongre üyeleri, New York bölgesinde benzer vakalar için cezaların üç yılla yirmi yıl hapis cezası arasında değişebildiği yönünde bulgulara vurgu yaparak özellikle, "açıklanamayacak şekilde yaygın" bir hüküm uyumsuzluğundan

dem vurdular. Yargıç Frankel'in önerdiği gibi yasa ABD Ceza Komisyonu'nun kurulmasını sağladı. Komisyonun başlıca görevi çok netti: zorunlu olması amaçlanan ve cezai hükümler için kısıtlı bir aralık tesis edecek bir hüküm verme kılavuzu hazırlamak.

Komisyon, bir sonraki yıl genel olarak on bin gerçek vakanın analizinde benzer suçlar için verilen ortalama hükümlere dayanan bu kılavuzu hazırladı. Süreçte büyük rol üstlenen Yüksek Mahkeme Yargıcı Stephen Breyer, komisyonun içinde giderilemeyen anlaşmazlığa dikkat çekerek geçmiş pratiklerin kullanılmasını savundu. "Komisyon neden sadece geçmişi almak yerine oturup akıl yoluyla hareket edemedi? Kısa cevap şu: Edemedik. Edemedik çünkü ortada aksi yönleri işaret eden çok sayıda son derece iyi argüman var... Var olan bütün suçları ceza hak ediş sırasına sokmaya çalışın... Sonra arkadaşlarınızdan sonuçları toplayın ve uyup uymadıkların bakın. Ben size uymayacaklarını söyleyebilirim."

Kılavuza göre yargıçlar, hüküm verirken iki faktörü göz önüne almalıydılar: Suç ve sanığın suç geçmişi. Suçlar ciddiyetlerine bağlı olarak kırk üç "saldırı seviyesi"nden birine dâhil ediliyordu. Sanığın suç geçmişi, ağırlıklı olarak daha önce giydiği hükümlerin sayısına ve ciddiyetine dayanıyordu. Suç ve suç geçmişi bir araya getirildikten sonra, kılavuz üst sınırın alt sınırı altı ay ya da %25 oranında aşabildiği, nispeten dar bir hüküm aralığı sunuyordu. Yargıçların, ağırlaştırıcı ya da hafifletici olarak gördükleri şartlara referansla bu aralığın tamamen dışına çıkmaları da mümkündü ama bu durumun bir temyiz mahkemesi tarafından tasdiklenmesi gerekiyordu.

Kılavuz zorunlu olmakla birlikte tamamen katı da değildi. Yargıç Frankel'in isteğini karşılamaya yaklaşmıyordu bile. Yargıçlara manevra yapabilecekleri hatırı sayılır bir alan bırakıyordu. Bununla birlikte çok çeşitli yöntemler kullanan ve belirli tarihsel dönemlere odaklanan birçok çalışma da aynı sonuca ulaştı: Kılavuz gürültüyü kesiyordu. Daha teknik bir ifadeyle, "hükmü

veren yargıcın kimliğine atfedilebilecek net hüküm çeşitliliklerini azaltmıştı."

En kapsamlı çalışma komisyonun kendisinden geldi. 1985 yılına (kılavuzun yürürlüğe girmesinden öncesine) ait banka soygunu, kokain ve eroin dağıtımı, bankada zimmetine para geçirme cezalarını 19 Ocak 1989 ve 30 Eylül 1990 arasındaki hükümlerle karşılaştırdı. Sanıklar kılavuzda hükümle alakalı kabul edilen faktörlere uygun olarak eşleştirildi. Ceza Reformu Yasası'nın yürürlüğe girmesinden sonraki dönemde, her suç için yargıçlar arasındaki farklılıklar hayli azalmıştı.

Bir başka çalışmaya göre 1986 ve 1987 yıllarında yargıçlar arasında beklenen hüküm süresi farkı, %17 veya 4,9 aydı. Bu sayı 1988 ve 1993 yılları arasında %11 ya da 3,9 aya düştü. Farklı dönemleri ele alan bağımsız bir çalışma da benzer dosya yüküne sahip yargıçlar arasındaki ortalama hüküm sürelerindeki farklar olarak tanımlanan yargıçlar arası uyuşmazlıkların azaltılmasında benzer bir başarı saptadı.

Kılavuz, bu bulgulara rağmen eleştiri yağmuruna tutuldu. Birçok yargıcın da aralarında olduğu bazı insanlar, kimi hükümlerin aşırı katı olduğunu düşünüyordu. Bu, gürültüyle değil ön yargıyla alakalı bir noktaydı. Bizim konumuz açısından daha ilginç olan ve çok sayıda yargıçtan gelen itiraz ise kılavuzun, yargıçların vakaya özel hususları yeterince dikkate almasını önlemesi nedeniyle son derece adaletsiz olduğuydu. Gürültüyü azaltmanın bedeli kabul edilemeyecek boyutta mekanik kararlar almaktı. Yale Üniversitesi'nden hukuk profesörü Kate Stith ve Federal Yargıç José Cabranes, "körlüğe değil içgörüye, hakkaniyete ihtiyaç olduğu"nu ve bunun "ancak bireysel vakanın çok katmanlılığını hesaba katan bir yargıyla mümkün olabileceğini" yazdılar.

Bu itiraz, kılavuza kimi hukuka, diğerleri politikaya dayalı şiddetli meydan okumalara yol açtı. Bu meydan okumalar burada özetlenen tartışmayla tamamen alakasız teknik nedenlerden

ötürü Yüksek Mahkeme'nin 2005 yılında kılavuzu lağvetmesine kadar amacına ulaşamadı. Mahkemenin kararının sonucu olarak kılavuz sadece tavsiye mahiyetine dönüştü. Yüksek Mahkeme'nin kararı sonrasında pek çok federal yargıç dikkate değer bir şekilde mutlu oldu. Yargıçların %75'i, bu mecburi rejimin tavsiye niteliğinde kalmasını tercih ederken, sadece %3'lük bir oran daha iyi olduğunu düşünüyordu.

Kılavuzu zorunludan tavsiye niteliğine değiştirmenin etkilerini neler oldu? Harvard Üniversitesi hukuk profesörü Crystal Yang, bu soruyu bir deneyle ya da bir anketle değil, gerçek hükümlerden oluşan ve yaklaşık dört yüz bin sanığı kapsayan çok büyük bir veri seti üzerinden soruşturdu. Yang'ın merkezî bulgusunun, çoklu ölçümlerle, yargıçlar arası uyumsuzlukların 2005 sonrası hatırı sayılır derecede arttığı oldu. Kılavuzun zorunlu olduğu dönemde nispeten sert bir yargıcın hüküm verdiği sanıklar, ortalama bir yargıcınkine göre 2,8 ay daha uzun hapis cezasına çarptırılıyordu. Kılavuzun tavsiye niteliğine dönüştürülmesiyle uyuşmazlık ikiye katlandı. Yang, Frankel'in kırk yıl önceki görüşlerine çok yakın bir ifadeyle, "bulgularının büyük hakkaniyet endişeleri uyandırdığını çünkü davaya verilen yargıcın kimliğinin benzer suçlardan hüküm giyen benzer suçluların farklı muameleler görmesine yabana atılmayacak bir katkı sağladığı"nı yazdı.

Kılavuzun tavsiye niteliğine dönmesinin ardından yargıçlar cezai hükümlerini daha ziyade kişisel değerlerine dayanarak vermeye başladılar. Zorunlu kılavuz gürültü kadar ön yargıyı da azaltmıştı. Yüksek Mahkeme'nin kararının ardından Afrikalı Amerikalı sanıklar ile aynı suçlardan hüküm giyen beyazların aldıkları cezalar arasındaki uyuşmazlıkta önemli bir artış oluştu. Yine kadın yargıçların artan takdir yetkilerini müsamahadan yana kullanma eğilimleri erkek yargıçlardan daha fazla arttı. Benzer bir durum, Demokrat başkanların atadığı yargıçlar için de geçerliydi.

Frankel'in 2002 yılında ölümünden üç yıl sonra zorunlu kılavuzun lağvedilmesi, yargıcın kâbusuna yakın bir şeye dönüşe yol açtı: düzenden yoksun hukuk.

Yargıç Frankel'in cezai hüküm kılavuzu için verdiği mücadelenin hikâyesi, bu kitapta ele alacağımız kilit konuların pek çoğu ile ilgili kısa bir fikir veriyor. Öncelikle yargıya varmak zordur çünkü dünya çok katmanlı ve belirsiz bir yerdir. Bu çok katmanlılık, hukuki yargılamada daha barizdir ama profesyonel yargı gerektiren durumların çoğu için de geçerlidir. Geniş anlamda bu durumlar doktorların, hemşirelerin, avukatların, mühendislerin, mimarların, Hollywood yapımcılarının, personel alım komisyonlarının, yayınevlerinin, kurumsal yöneticilerin ve spor takımlarının yönetimlerinin aldığı kararları da kapsar. Yargının işin içine girdiği her yerde anlaşmazlık kaçınılmazdır.

İkinci olarak bu anlaşmazlıkların kapsamı, beklediğimizden çok daha büyüktür. Hukuki takdir yetkisi prensibine çok az insan karşı çıkmakla birlikte bu yetkinin ürettiği farklılıkların boyutunu kınamayanlar yok denecek kadar azdır. *Sistem gürültüsü*, yani ideal olarak birbirinin eşi olması gereken yargılarda istenmeyen bir çeşitlilik olması, yaygın adaletsizliğe, çok yüksek ekonomik maliyetlere ve çok çeşitli hatalara yol açabilir.

Üçüncü olarak gürültü azaltılabilir. Frankel'in savunduğu ve ABD Ceza Komisyonu'nun kurallar ve kılavuzla yürürlüğe koyduğu yaklaşım, gürültüyü başarıyla azaltan çok sayıdaki yaklaşımdan sadece biridir. Diğer yaklaşımlar başka yargı tiplerine daha uygundur. Gürültüyü azaltmak için benimsenen bazı yöntemler eş zamanlı olarak ön yargıyı da azaltabilir.

Dördüncü olarak gürültüyü azaltma çabaları, sıklıkla itirazlara neden olur ve bu süreçte ciddi zorluklarla karşılaşır. Bu meseleler de ele alınmalıdır, yoksa gürültüyle mücadele başarısız olur.

Cezai Hükümde Gürültüden Bahsederken

"Deneyler yargıçların benzer vakalar için vardıkları hükümler arasında büyük farklılıklar gösteriyor. Bu çeşitlilik adil olamaz. Bir sanığın alacağı ceza o davanın hangi yargıca verileceğine bağlı olmamalıdır."

"Cezai hükümler yargıcın oturum sırasındaki ruh hâline ya da dışarıdaki hava sıcaklığına bağlı olmamalıdır."

"Kılavuzlar bu meseleyi ele almanın bir yoludur. Ancak adaleti ve doğruluğu sağlamak için gerekli olabilecek yargısal takdir yetkisini kısıtladıkları için, pek çok insan kılavuzlardan hoşlanmaz. Sonuçta her dava benzersiz değil midir?"

BÖLÜM 2

Gürültülü Bir Sistem

Gürültüyle ilk karşılaşmamız ve bu konuya ilgimizi tetikleyen ilk şey, adalet sistemiyle ters düşmemiz gibi çarpıcı bir olay değildi. Aslında, ilk karşılaşmamız aramızdan iki kişinin bağlantılı olduğu danışmanlık şirketine işveren bir sigorta şirketinin dâhil olduğu bir tür kazaydı.

Elbette sigorta konusu herkesin ilgi alanına girmeyebilir. Öte yandan bulgularımız, gürültülü kararlardan çok şey kaybetme riski olan kâr amaçlı örgütlerde, gürültü sorununun büyük olduğunu gösteriyor. Sigorta şirketiyle yaşadığımız tecrübe, sorunun neden genellikle gözden kaçtığını ve bu konuda ne yapılabileceğini açıklamaya yardımcı olacaktır.

Sigorta şirketinin üst düzey yöneticileri, şirket adına önemli finansal kararlar alan kişilerin yargılarında tutarlılığı artırmak –dolayısıyla gürültüyü azaltmak– için harcanacak çabanın potansiyel değerini tartıyorlardı. Tutarlılığın önemli olduğu konusunda herkes hemfikirdi. Herkes ayrıca bu yargıların, gayriresmî ve kısmen öznel oldukları için, hiçbir zaman tamamen tutarlı olamayacağı konusunda da hemfikirdi. Bir miktar gürültü kaçınılmazdı.

Konu gürültünün boyutuna gelince, anlaşmazlık baş gösterdi. Üst düzey yöneticiler gürültünün şirketleri için önemli bir sorun olabileceğinden şüpheleniyorlardı. Ancak haklarını vermek gerek, soruyu bizim *gürültü denetimi* dediğimiz basit bir deneyle halletmeyi kabul ettiler. Sonuç onları şaşırtmıştı. Aynı zamanda gürültü sorununun kusursuz bir resmi de ortaya çıkmıştı.

Gürültü Çıkaran Bir Piyango

Bütün büyük şirketlerde pek çok profesyonele şirketi bağlayıcı kararlar alma yetkisi verilir. Örneğin bu sigorta şirketi, dolandırıcılık ya da çalışanların gerçekleştirdiği yasa dışı gizli operasyonlara bağlı kayıplara karşı bir bankayı sigortalamak gibi, finansal riskler için prim hesaplayan sayısız risk değerlendirme uzmanı çalıştırıyor. Ayrıca bünyesinde görevi gelecekteki hasar taleplerinin maliyetini öngörmek ve anlaşmazlık oluşması hâlinde, hasar talebinde bulunan kişilerle pazarlık etmek olan sigorta eksperleri de var.

Şirketin bütün büyük şubelerinde birkaç kalifiye risk değerlendirme uzmanı bulunuyor. Bir fiyat teklifi talep edildiğinde, o anda kim müsaitse hazırlama görevini o üstleniyor. Buna söz konusu risk değerlendirme uzmanı kurayla seçiliyor da diyebiliriz.

Verilen fiyatın tam değerinin şirket açısından önemi çok büyük. Teklif kabul edilirse yüksek bir prim avantajlı olur ama yüksek bir prim işi bir rakibe kaptırmaya neden olabilir. Düşük bir primin kabul edilme olasılığı daha yüksek olsa da şirket için daha az avantaj demek. Her risk için tam doğru –ne çok yüksek ne çok alçak– olacak makul bir fiyat var ve büyük bir profesyoneller grubunun ortalama kararının bu makul sayıya çok uzak olmama ihtimali çok yüksek. Bu sayıdan yüksek ya da düşük olan teklifler çok pahalıya mal olabilir; gürültülü yargıların değişkenliğinin bilançonun son satırına zararı bu şekilde olur.

Sigorta eksperlerinin işi, şirketin finansal durumunu da etkiler. Örneğin endüstriyel bir kaza sonucu sağ eli kalıcı olarak kullanılamaz hâle gelmiş bir işçi adına bir hasar talebinde bulunulduğunu varsayalım. Bu hasar talebi için, tıpkı başlangıçta fiyat teklifi yapan risk değerlendirme uzmanının görevlendirildiği gibi, o anda tesadüfen müsait olan bir sigorta eksperi görevlendirilsin. Sigorta eksperi, vakanın olgularını toplar ve şirkete nihai maliyet konusunda bir tahmin hazırlar. Aynı eksper daha sonra hasar talebinde bulunan kişi poliçede vadedilen menfaatleri alırken, şirketi aşırı bir ödeme yapmaktan korumak için sigortalının temsilcisiyle pazarlık görevini de üstlenir.

İlk tahmin önemlidir çünkü sigorta eksperine sigortalıyla gelecek pazarlıklar için örtülü bir hedef belirler. Sigorta şirketi ayrıca yasal olarak her hasar talebinin öngörülen maliyetini rezerve etmekle de (başka bir deyişle, ödeyebilecek nakdi bulundurmakla da) yükümlüdür. Burada bir kez daha şirket açısından makul bir değer söz konusudur. Diğer tarafta sigortalıyı temsil eden ve teklifin yetersiz gelmesi durumunda mahkemeye gitmeyi tercih edebilecek bir avukat olduğu için, uzlaşma garanti değildir. Öte yandan aşırı cömert bir rezerv, sigorta eksperine karşı tarafın uçarı taleplerini kabul etmesine neden olacak aşırı bir esneklik verebilir. Sigorta eksperinin yargısı, şirket için büyük önem taşır; hatta sigortalı için bu önem daha da büyüktür.

Piyango kelimesini risk değerlendirme uzmanının ya da sigorta eksperinin seçilmesinde şansın rolünü vurgulamak için kullandık. Şirketin normal işleyişinde, her vakaya tek bir profesyonel atanır ve onun yerine başka bir çalışma arkadaşı seçilmiş olsa neler olacağını kimse bilemez.

Piyangoların kendilerine ait bir yeri var ve adaletsiz olmamalılar. Kabul edilebilir piyangolar, bazı üniversitelerde kurslar gibi "iyi" şeyleri ve ya da zorunlu askerlik gibi "kötü" şeyleri tahsis etmek için kullanılır. Bir amaca hizmet ederler. Ancak bahset-

tiğimiz yargı piyangoları bir şeyi tahsis etmez; sadece belirsizlik üretir. Risk değerlendirme uzmanlarının gürültüsüz olduğu ve optimal primler ayarladığı ama daha sonra müşterinin göreceği teklif rakamını değiştirmek için şans aracının müdahale ettiği bir sigorta şirketi düşünün. Elbette, böyle bir piyangoyu haklı göstermek mümkün olmaz. Aynı şekilde sonucun profesyonel bir karar almak için rastgele seçilmiş bir insanın kimliğine bağlı olduğu bir sistemi de haklı gösterecek bir gerekçe yoktur.

Gürültü Denetimleri Sistem Gürültüsünü Açığa Çıkarır

Bir cezai hükmü belirlemek için belli bir yargıcı ya da bir takımı temsil etmek için tek bir nişancı seçen piyango, değişkenlik yaratır ama bu değişkenlik gözle görülmez. Gürültü denetimi –cezai hükümler konusunda federal yargıçlar üstünde gerçekleştirilen denetim gibi– gürültüyü açığa çıkarmanın bir yoludur. Böyle bir denetimde aynı vaka birden fazla birey tarafından değerlendirilir ve tepkilerinin çeşitliliği gözler önüne serilir.

Risk değerlendirme uzmanlarının ve sigorta eksperlerinin yargıları, bu deney için özellikle uygun çünkü kararları yazılı bilgiye dayalı. Gürültü denetimine hazırlanmak için, şirketin üst düzey yöneticileri her grup için (risk değerlendirme uzmanları ve sigorta eksperleri) beş temsilî vakanın detaylı tanımını oluşturdu. Çalışanların her birinden, bağımsız çalışarak iki ya da üç vakayı değerlendirmeleri istendi. Çalışmanın amacının yargılarının değişkenliği olduğunu incelemek olduğundan bahsedilmedi.

Okumaya devam etmeden önce, aşağıdaki sorulara kendi cevabınızı düşünmek isteyebilirsiniz. İyi çalışan bir sigorta şirketinde rastgele iki kalifiye risk değerlendirme uzmanı ya da sigorta eksperi seçseniz aynı vaka için tahminlerinin ne kadar farklı

olmasını beklersiniz? Daha net bir ifadeyle iki tahmin arasındaki fark, ortalamalarının bir yüzdesi olarak ne olurdu?

Şirketteki birçok üst düzey yöneticiye cevaplarını sorduk ve sonraki yıllarda farklı mesleklerden çok çeşitli insandan tahminler aldık. Şaşırtıcı bir şekilde bir cevap diğerlerinden bariz bir şekilde daha popülerdi. Sigorta şirketinin çoğu üst düzey yöneticisi %10 ya da daha az tahmininde bulundu. Çok çeşitli sektörlerden 828 CEO ve kıdemli yöneticiye benzer uzman yargılarında ne kadar sapma beklediklerini sorduğumuzda, %10 yine medyan ve en sık verilen cevaptı (ikinci en popüler cevap ise %15'ti). %10'luk bir fark, örneğin, iki risk değerlendirme uzmanından biri 9.500 dolarda karar kılarken, diğerinin 10.500 dolarlık bir teklif yapması anlamına geliyor. Önemsiz bir fark değil ama bir organizasyonun tolere etmesi beklenen bir fark.

Bizim gürültü denetimimiz çok daha büyük farklar buldu. Bizim ölçümümüzle, sigortalamadaki medyan fark %55 yani şirketin üst düzey yöneticileri dâhil pek çok insanın beklediğinin yaklaşık beş katıydı. Bu sonuç şu anlama geliyor: Örnek olarak bir risk değerlendirme uzmanı 9.500 dolarlık bir teklif verdiğinde, diğeri 10.500 dolar değil, 16.700 dolar hesaplıyor. Siporta eksperleri içinse medyan oran %43 idi. Bu sonuçların medyan olduğunun altını çizmek isteriz: Vaka çiftlerinin yarısında iki yargı arasındaki fark daha da büyüktü.

Gürültü denetiminin sonuçlarını raporladığımız üst düzey yöneticiler gürültünün büyüklüğünün, pahalı bir sorunu temsil ettiğini anlamakta gecikmediler. Kıdemli bir üst düzey yönetici sigortalamada şirketin yıllık gürültü maliyetinin –hem aşırı yüksek teklifler yüzünden kaybedilen işleri hem de düşük fiyatlı kontratların neden olduğu kayıpları hesaba katarak– yüzlerce milyon dolar olduğu tahmininde bulundu.

Ne kadar hata (ya da ön yargı) olduğunu kimse net olarak söyleyemezdi çünkü her vaka için makul değeri kimse kesin olarak

bilemezdi. Bununla birlikte hedef tahtasının arkasındaki dağılımı ölçmek ve değişkenliğin bir sorun olduğunu fark etmek için kimsenin hedefi görmesine gerek yoktu. Veriler, bir müşterinin ödemesi istenecek fiyatın, rahatsız edici bir dereceye kadar, o işlemle ilgilenecek çalışanı seçen piyangoya bağlı olduğunu göstermişti. Müşteriler, kendi rızaları olmadan böyle bir piyangoya katıldıklarını duymaktan en hafif tabirle pek hoşnut kalmazdı Daha genel anlamda organizasyonlarla muhatap olan insanlar, karşılarında tutarlı yargılara varan bir sistem bulmayı beklerler. Sistem gürültüsüyle karşılaşmayı beklemezler.

İstenen Çeşitliliğe Karşı İstenmeyen Değişkenlik

Sistem gürültüsünün tanımlayıcı bir özelliği *istenmiyor* olmasıdır; bu anlamda burada, yargılarda değişkenliğin her zaman istenmeyen bir şey olmadığının altını da çizmeliyiz.

Tercih ve zevk meselelerini düşünün. On film eleştirmeni aynı filmi izlediğinde, on şarap tadımcısı aynı şarabı puanladığında ya da on kişi aynı romanı okuduğunda hepsinin aynı görüşe sahip olmasını beklemeyiz. Zevkte çeşitlilik hoş karşılanan ve tamamen beklenen bir şeydir. Hiç kimse herkesin aynı şeyleri beğenip aynı şeylerden hoşlanmadığı bir dünyada yaşamak istemez. (Neredeyse hiç kimse.) Ancak kişisel zevk profesyonel yargıyla karıştırıldığı zaman, zevklerin çeşitliliği hataların açıklanmasına yardımcı olabilir. Bir film yapımcısı senaryoyu şahsen sevdiği için sıra dışı bir projeye (diyelim, çevirmeli telefonların yükselişi ve çöküşü hakkında olsun) girmeye karar vermişse, filmi başka kimse sevmezse çok büyük bir hata yapmış olabilir.

En iyi yargının ödüllendirileceği rekabetçi bir durumda yargılarda çeşitlilik yine beklenir ve bu hoş karşılanır. Birden fazla şirket (ya da aynı organizasyonun içinde birçok ekip) aynı müşteri sorununa yenilikçi çözümler üretmek için rekabete girdiğinde,

aynı yaklaşıma odaklanmalarını istemeyiz. Bu durum, bir aşının geliştirilmesi örneğinde olduğu gibi bilimsel bir soruna birden fazla araştırma ekibi yoğunlaştığında da geçerlidir: Ekiplerin her birinin olaya farklı açılardan bakmalarını isteriz. Bazen tahminciler bile rakip oyuncular gibi davranırlar. Başka hiç kimsenin beklemediği bir gerilemeyi doğru olarak öngören bir analistin, ün kazanacağı kesindir; konsensüsten hiç sapmayan analist ise gölgede kalır. Bu tür ortamlarda fikirlerde ve yargılarda farklılık yine hoş karşılanır çünkü farklılık sadece ilk adımdır. İkinci bir safhada bu yargıların sonuçları birbiriyle çarpıştırılacak ve en iyi olan kazanacaktır. Piyasalarda, tıpkı doğada olduğu gibi, farklılık olmadan ayıklanma olamaz.

Zevk ve rekabetçi ortam meseleleri ilginç yargı sorunları oluşturur. Ama bizim ilgi odağımız farklılığın istenmediği yargılardır. Sistem gürültüsü sistemlerin sorunudur ve sistemler piyasa değil, organizasyonlardır. Borsacılar bir hissenin değeri üzerinde farklı değerlendirmeler yaptığında, kimi para kazanır kimi kazanmaz. Piyasaları görüş ayrılıkları yaratır. Ancak söz konusu borsacılardan biri o değerlendirmeyi firması adına yapmak için rastgele seçildiğinde ve aynı şirketteki çalışma arkadaşlarının çok farklı değerlendirmeler ürettiğini öğrenmemiz durumunda şirket sistem gürültüsüyle karşı karşıyadır ve bu bir sorundur.

Bulgularımızı bir portföy yönetimi şirketinin kıdemli yöneticilerine sunduğumuz ve onları kendi açıklayıcı gürültü denetimlerini yapmaya davet ettiğimizde, meselenin zarif bir izahı ortaya çıktı. Şirketteki kırk iki tecrübeli yatırımcıdan bir hisse senedinin makul değerini (yatırımcıların almaya ya da satmaya kayıtsız kalacağı fiyatı) tahmin etmelerini istediler. Yatırımcılar analizlerini şirketin bir sayfalık tarifine dayandırdılar. Söz konusu veriler son üç yılın basitleştirilmiş kâr ve zarar tablosunu, bilançolarını ve nakit akış beyanlarını ve sonraki iki sene için tahminleri içeriyordu. Sigorta şirketindeki gibi ölçülen medyan gürültü %41'di. Aynı şirketin

aynı değerlendirme yöntemlerini kullanan yatırımcıları arasında bu kadar büyük farklar iyi haber olamaz.

Bu portföy yönetimi şirketinde, ceza yargı sisteminde ve daha önce ele aldığımız sigorta şirketinde olduğu gibi, yargıya varacak kişinin eşit kalifikasyonda bireyler arasından rastgele seçildiği bütün durumlarda gürültü sorunu yaşanır. Sistem gürültüsü pek çok örgüte zarar verir: Genellikle bir hastanede sizi hangi doktorun göreceğine, mahkemede davanıza hangi yargıcın bakacağına, başvurunuzu hangi patent uzmanının gözden geçireceğine, şikâyetinizi hangi müşteri hizmetleri temsilcisinin dinleyeceğine rastgele bir görevlendirme süreci karar verir. Bu yargılarda istenmeyen farklılıklar para kaybı ve dizginsiz haksızlığın da aralarında olduğu ciddi sorunlara rol açabilir.

Yargılarda istenmeyen farklılıklarla ilgili sık rastlanan bir yanlış algı, rastgele hataların sözüm ona birbirini dengelemesi nedeniyle önemli olmadıklarıdır. Elbette, aynı vakayla ilgili bir yargıdaki olumlu ve olumsuz hatalar birbirlerini dengeleme eğilimindedir; bu özelliğin gürültüyü azaltmak için nasıl kullanılacağını detaylı olarak ele alacağız. Ancak gürültülü sistemler aynı vakadan çok sayıda yargı üretmezler. Farklı vakalarda gürültülü yargılar üretirler. Bir sigorta poliçesinde prim çok yüksek, bir diğerinde çok düşükse, fiyatlandırma ortalamada doğru görünebilir ama sigorta şirketi maliyeti yüksek iki hata yapmış demektir. Her ikisi de beş yıl hapis cezasına çarptırılması gereken iki suçludan biri üç, biri yedi yıl ceza alırsa, adalet ortalamada sağlanmış olmaz. Gürültülü sistemlerde hatalar birbirini dengelemez. Üst üste binerler.

Anlaşma Yanılsaması

Profesyonel yargıda gürültüyü analiz etmek için birkaç on yıllık literatüre bakmamız gerekir. Bu literatürün varlığından haberdar olduğumuz için sigorta şirketinin gürültü denetimi bizi şaşırtmadı.

Bizi şaşırtan, bulgularımızı raporladığımız üst düzey yöneticilerin tepkisi oldu: Şirketteki hiç kimse gözlemlediğimiz gürültü miktarını beklemiyordu. Denetimin geçerliliğini sorgulayan, gürültü miktarının kabul edilebilir olduğunu iddia eden olmadı. Bununla birlikte gürültü problemi –ve bunun büyük maliyeti– organizasyon için yeni bir sorun gibi görünüyordu. Gürültü bodrum katındaki bir sızıntıydı. Kabul edilebilir görüldüğü için değil, fark edilmediği için hoş görülmüştü.

Bu nasıl olabilmişti? Aynı ofiste aynı rolü oynayan profesyoneller farkında olmadan birbirlerinden nasıl bu kadar farklı olabilmişlerdi? Üst düzey yöneticiler şirketlerinin performansı ve itibarı için önemli bir tehdit olduğunu anladıkları bu gözlemi nasıl yapamamışlardı? Sistem gürültüsü sorununun örgütlerde sık sık gözden kaçırıldığını ve gürültüye karşı bu yaygın dikkatsizliğin, gürültünün yaygınlığı kadar ilginç olduğunu fark ettik. Gürültü denetimleri, saygı duyulan profesyonellerin –ve onları çalıştıran organizasyonların– aslında günlük profesyonel yargılarında fikir ayrılığı yaşarken bir *anlaşma yanılsamasını* koruduklarını düşündürüyordu.

Anlaşma yanılsamasının nasıl doğduğunu anlamaya başlamak için, kendinizi normal bir iş gününde bir risk değerlendirme uzmanının yerine koyun. Beş yıldan daha fazla tecrübeniz var, çalışma arkadaşlarınızın sizin hakkınızda iyi düşünceler taşıdığını biliyorsunuz ve siz de onlara saygı ve sevgi duyuyorsunuz. İşinizde iyi olduğunuzu biliyorsunuz. Bir finans şirketinin karşı karşıya olduğu çok katmanlı, karmaşık riskleri enine boyuna analiz ettikten sonra, iki yüz bin dolarlık bir primin uygun olduğu sonucuna vardınız. Sorun karmaşık ama haftanın her günü çözdüklerinizden çok farklı değil.

Şimdi, ofiste çalışma arkadaşlarınıza da aynı bilginin verildiğini ve onların da aynı riski değerlendirdiklerini hayal edin. En az yarısının iki yüz elli beş bin dolardan daha yüksek ya da yüz

kırk beş bin dolardan daha düşük bir prim belirlediğine inanabilir misiniz? Kabul etmesi zor bir düşünce. Hatta gürültü denetimini duyan ve geçerliliğini kabul eden risk değerlendirme uzmanlarının bile, denetim sonuçlarının kendileri için de geçerliği olduğuna hiçbir zaman gerçekten inanmadığından şüpheleniyoruz.

Çoğumuz, çoğu zaman, hiç sorgulamadan dünyanın böyle olduğu için böyle göründüğü inancıyla yaşarız. Bu inanç bir diğerine, yani "Diğer insanlar dünyayı çoğunlukla benim gördüğüm gibi görüyor." inancına sadece küçük bir adım uzaktır. *Naif gerçekçilik* olarak adlandırılan bu inançlar, diğer insanlarla paylaştığımız gerçeklik hissi için esastır. Bu inançları nadiren sorgularız. Her zaman etrafımızdaki dünyaya ait tek bir yoruma sahibizdir ve normalde bu yoruma akla yatkın alternatifler üretmeye çok çaba harcamayız. Bir yorum yeterlidir ve onu gerçek olarak tecrübe ederiz. Hayatı, gördüklerimizi görmenin alternatif yollarını hayal ederek yaşamayız.

Profesyonel yargılar örneğinde, başka insanların dünyayı bizim gibi gördüğü inancı her gün çok çeşitli şekillerde pekiştirilir. İlk olarak, çalışma arkadaşlarımızla ortak bir dil ve kararlarımızda önemli olması gereken faktörlerle ilgili bir dizi kural paylaşırız. Ayrıca bu kuralları çiğneyen yargıların saçmalığı konusunda iç rahatlatıcı bir şekilde hemfikir olma tecrübesine de sahibizdir. Meslektaşlarımızla zaman zaman yaşadığımız anlaşmazlıkları onların muhakeme hatası olarak görürüz. Üstünde anlaşılmış kurallarımızın müphem ve bazı olasılıkları ortadan kaldırmaya yeterli olmakla birlikte, belli bir vakaya ortak bir olumlu tepki belirlemek için yetersiz olduğunu fark etme ihtimalimiz çok azdır. Diğer çalışma arkadaşlarımızla, dünyayı aslında bizim gibi görmediklerini hiç fark etmeden geçinir gideriz.

Mülakat yaptığımız bir risk değerlendirme uzmanı, kendi departmanında kıdemli eleman olma tecrübesini anlattı: "Başlarda, vakalarımın yüzde yetmiş beşini şefimle tartışırdım... Birkaç sene

sonra, gerek kalmadı; artık uzman olarak görülüyordum... Zamanla yargılarımdan daha emin oldum." Pek çoğumuz gibi bu kişi de kendi yargısına güvenini onu uygulayarak geliştirmişti.

Bu sürecin psikolojisi çok açıktır. Öz güven, gittikçe artan bir akıcılık ve kolaylıkla varılan yargıların öznel tecrübesinden beslenir. Bu, kısmen yeni yargıların geçmişte benzer durumlarda varılan yargılara benzerliklerinden kaynaklanır. Bu risk değerlendirme uzmanı zaman içinde, geçmişteki kendisiyle aynı fikre varmayı öğrendikçe, yargılarına güveni de arttı. Çıraklık döneminin ardından başkalarıyla benzer düşünmeyi öğrendiğine, onlarla ne dereceye kadar hemfikir olduğunu kontrol ettiğine ve hatta uygulamalarının çalışma arkadaşlarınınkilerden uzaklaşmasını önlemeye çalıştığına dair hiçbir işaret sergilemiyordu.

Sigorta şirketi için anlaşma yanılsaması ancak gürültü denetiminin yapılmasıyla paramparça oldu. Şirketin liderleri gürültü sorunlarından nasıl bihaber kalmışlardı? Burada birkaç olası cevap var ancak pek çok ortamda önemli bir rol oynar gibi görünen cevap basitçe anlaşmazlığın verdiği rahatsızlıktır. Çoğu organizasyon konsensüsü ve uyumu uyumsuzluk ve çatışmaya yeğler. Yürürlükteki süreçler genellikle gerçek anlaşmazlıklara maruz kalma sıklığını asgari düzeye indirmek ve bu tür anlaşmazlıklar oluştuğunda onları açıklayarak etkilerini hafifletmek için tasarlanmış gibi görünüyorlar.

Minnesota Üniversitesi'nde psikoloji profesörü ve performans öngörüsü konusunda önde gelen araştırmacılardan Nathan Kuncel, bizimle bu sorunu çok iyi anlatan bir hikâye paylaştı. Kuncel bir okulun öğrenci kabul ofisinin karar sürecini gözden geçirmesine yardımcı oluyordu. Başvuru dosyasını önce bir kişi okuyor, daha sonra puanlamalarıyla birlikte ikinci bir okuyucuya aktarıyor, bu ikinci okuyucu da dosyayı puanlıyordu. Kuncel, bu kitap ilerledikçe daha anlaşılır olacak nedenlerden ötürü, ikinci okuyucuyu etkilememesi için ilk okuyucunun verdiği puanların gizlenmesinin daha

doğru olabileceğini belirtti. Buna karşılık okulun cevabı şu oldu: "Eskiden öyle yapardık ama bu durum çok fazla anlaşmazlığa yol açınca, mevcut sisteme geçtik." Bu okul çatışmadan kaçınmayı, en az doğru karar almak kadar önemli gören tek organizasyon değil.

Birçok şirketin sığındığı bir başka mekanizmayı, talihsiz yargıların otopsisini ele alalım. Otopsiler bir öğrenme mekanizması olarak faydalıdır ama eğer gerçekten bir yargının profesyonel normların çok dışına çıkması anlamında bir hata yapıldıysa, hatayı tartışmak zorlayıcı olmayacaktır. Uzmanlar yargının konsensüsün dışında kaldığına kolayca karar verirler. (Ayrıca nadir bir istisna olarak üstünü de çizebilirler.) Kötü yargıyı teşhis etmek iyi bir yargıyı teşhis etmekten daha kolaydır. Muazzam kötü hataları haykırmak ve kötü çalışma arkadaşlarını kenara itmek, profesyonellerin yaygın olarak kabul edilebilir yargılara varırken ne çok fikir ayrılığı yaşadıklarını fark etmelerine yardımcı olmayacaktır. Aksine kötü yargılar hakkında kolayca uzlaşmaya varmak, bir anlaşma yanılsamasının pekişmesine bile neden olabilir. Sistem gürültüsünün her yerde var olmasıyla ilgili gerçek ders asla alınmaz.

Sistem gürültüsünün ciddi bir sorun olduğu görüşümüzü paylaşmaya başladığınızı umuyoruz. Varlığı bir sürpriz değil; gürültü yargının gayriresmî doğasının bir sonucudur. Ancak bu kitap boyunca göreceğimiz gibi bir organizasyon durumu ciddiyetle ele aldığında, gözlenen gürültü miktarı neredeyse her zaman şok etkisi yaratır. Bizim çıkarımımız çok basittir: Yargının olduğu yerde gürültü vardır, hem de tahmin edeceğinizden çok daha fazla.

Sigorta Şirketinde Sistem Gürültüsünden Bahsederken

"Risk değerlendirme uzmanlarının ve sigorta eksperlerinin profesyonel yargılarının kalitesine güveniriz. Her vakayı bir uzmana

veririz ama bir başka uzmanın benzer bir yargı üreteceği yanılgısıyla hareket ederiz."

"Sistem gürültüsü sandığımızdan ya da kaldırabileceğimizden beş kat daha büyüktür. Gürültü denetimi olmasaydı bunu asla fark etmezdik. Gürültü denetimi anlaşma yanılsamasını yerle bir etmiştir."

"Sistem gürültüsü ciddi bir sorun: Bize milyonlarca dolara mal oluyor."

"Yargının olduğu yerde gürültü vardır; tahmin edeceğinizden çok daha fazla."

BÖLÜM 3

Tekil Kararlar

Şu ana kadar ele aldığımız vaka örnekleri, defaatle varılan yargıları içeriyordu. Hırsızlıktan hüküm giyen biri için doğru ceza nedir? Belli bir risk için doğru prim nedir? Her vaka bir anlamda benzersiz olsa da bu tür yargılar *tekrarlı kararlar*dır. Hastalara teşhis koyan doktorlar, şartlı tahliye davalarına bakan yargıçlar, başvuruları gözden geçiren kabul memurları, vergi formlarını hazırlayan muhasebeciler; bunların her biri tekrarlı kararlara birer örnektir.

Tekrarlı kararlarda gürültü bir önceki bölümde bahsettiklerimiz gibi gürültü denetimleriyle ortaya konabilir. Farklı profesyoneller benzer vakalarda karar aldığında, istenmeyen farklılığı tanımlamak ve ölçmek kolaydır. Gürültü fikrini *tekil kararlar* olarak adlandırdığımız yargı kategorisine uygulamak daha zor, hatta belki de imkânsız görünür.

Örneğin dünyanın 2014 yılında karşı karşıya kaldığı krizi ele alalım. Batı Afrika'da çok sayıda insan Ebola'dan hayatını kaybediyordu. Dünya kendi içinde bağlantılı olduğu için, enfeksiyonların bütün dünyaya hızla yayılacağı, Avrupa ve Kuzey Amerika'yı özellikle kötü sarsacağı tahminleri yürütüldü. Amerika Birleşik

Devletleri'nde virüsten etkilenen bölgelerden hava yolculuğunu durdurmak ve sınırları kapatmak için agresif adımlar atmaya dönük ısrarcı çağrılar yapıldı. Bu yönde ilerleme konusunda politik baskı çok yoğundu ve öne çıkan bilgili insanlar bu adımlardan yanaydı.

Başkan Barack Obama başkanlık döneminin, bir benzerine daha önce rastlamadığı ve daha sonra da rastlamayacağı en zor kararlarından biriyle karşı karşıyaydı. Sınırları kapatmamayı seçti. Bunun yerine Batı Afrika'ya –sağlık çalışanları ve askerlerden oluşan– üç bin kişi gönderdi. Her zaman birlikte uyumlu hareket etmeyen çok çeşitli, uluslararası bir milletler koalisyonuna öncülük etmenin yanı sıra kaynaklarını ve uzmanlıklarını sorunun kaynağında üstesinden gelmek için kullandı.

Tekrarlıya Karşı Tekil

Başkanın Ebola'ya tepkisi gibi bir kez verilen kararlar aynı birey ya da takım tarafından tekrarlı olarak alınmadıkları, el altında kullanıma hazır bir tepki bulunmadığı ve hakikaten benzersiz özelliklere sahip oldukları için tekildirler. Başkan Obama ve ekibinin Ebola ile uğraşırken referans alabileceği daha eski bir örnek yoktu. Önemli politik kararlar genellikle tekil kararlara çok iyi örnektir; tıpkı ordu komutanlarının tarihî önem taşıyan tercihleri gibi.

Kişisel hayatta bir işi seçerken, bir ev alırken ya da evlenme teklif ederken aldığınız kararların benzer özellikleri vardır. Bu ilk işiniz, eviniz ya da evliliğiniz olmasa ve sizden önce sayısız insan aynı kararları almış olsa bile, karar size benzersiz görünür. İş hayatında, şirketlerin başındaki insanlar sık sık onlara benzersiz gibi görünen kararlar almak durumunda kalırlar: sektörde gidişatı değiştirme potansiyeline sahip bir yeniliği piyasaya sürüp sürmemek, bir salgın sırasında ne kadar kapanılacağı, yabancı bir ülkede ofis açıp açmamak ya da onları denetim altına almaya çalışan bir hükûmete teslim olup olmamak gibi.

Tartışmaya açık olsa da tekil ve tekrarlı kararlar arasında bir kategori farkı değil, bir süreklilik vardır. Risk değerlendirme uzmanları, onlara çok olağan dışı görünen davalarla uğraşabilirler. Tam tersine hayatınızda dördüncü kez bir ev satın alıyorsanız ev satın almayı tekrarlı bir karar olarak görmeye başlamış olabilirsiniz. Ama bu uç örnekler, açık bir şekilde farkın anlamlı olduğunu düşündürür. Savaşa girmek başka bir şeydir, yıllık bütçeyi gözden geçirmek bambaşka.

Tekil Kararlarda Gürültü

Tekil kararlar geleneksel olarak büyük organizasyonlarda, aralarında yer değiştirebilen çalışanların rutin olarak aldığı tekrarlı kararlardan ayrı bir yere konur. Sosyal bilimciler tekrarlı kararları ele alırken, riskin yüksek olduğu tekil kararlar tarihçilerin ve yönetim gurularının uzmanlık alanı olmuştur. İki karar türüne yaklaşım hayli farklıdır. Tekrarlı kararların analizleri, sosyal bilimcilerin örüntüleri ayırt etmek, düzenliliği teşhis etmek ve doğruluğu ölçmek için pek çok benzer kararı değerlendirmesi nedeniyle, sıklıkla istatistiksel bir boyut alır. Öte yandan tekil kararların tartışılması tipik olarak nedensel bir bakış açısı benimser; bu tartışmalar geriye dönüp bakılarak yürütülür ve olanların nedenlerini saptamaya odaklıdırlar. Yönetsel başarı ve başarısızlıklara ilişkin vaka incelemeleri gibi tarihî analizler, özünde benzersiz olan bir yargıya nasıl varıldığını anlamayı hedef edinir.

Tekil kararların doğası, gürültünün incelenmesi açısından önemli bir soruyu gündeme getirir. Gürültüyü aynı sorunun yargılarındaki istenmeyen farklılıklar olarak tanımlamıştık. Tekil kararlar hiçbir zaman aynı şekilde tekrarlanmadığı için bu tanım onlara uymaz. Sonuçta tarih bir kez yaşanır. Obama'nın 2014 yılında Batı Afrika'ya sağlık çalışanı ve asker gönderme kararını, diğer Amerikan başkanlarının aynı dönemde aynı sorunu ele almak için verdikleri

kararlarla hiçbir zaman mukayese edemeyeceksiniz (sadece fikir yürütebilirsiniz). O özel kişiyle evlenme kararınızı sizin gibi diğer insanların kararlarıyla karşılaştırmayı kabul edebilirsiniz ama o karşılaştırma size farklı risk değerlendirme uzmanlarının aynı vakada verdiği teklifler arasındaki karşılaştırmamız kadar alakalı gelmeyecektir. Siz ve eşiniz, benzersizsiniz. Tekil kararlarda gürültünün varlığını gözlemlemenin direkt bir yolu yok.

Bununla birlikte tekil kararlar, tekrarlı kararlarda gürültü üreten faktörlerden tamamen arınmış değildir. Atış poligonunda (gürültülü takım) Takım C'nin nişancıları tüfeklerinin nişangâhını farklı yönlerde ayarlıyor ya da sadece elleri titriyor olabilir. Sadece ilk nişancısını gözlemleseydik takımın ne kadar gürültülü olduğu konusunda bir fikrimiz olmazdı ama gürültü kaynakları hâlâ var olurdu. Benzer şekilde tekil bir karar aldığınız zaman, başka bir karar alıcının, sizinle aynı beceriye, aynı hedef ve değerlere sahip bile olsa, aynı olgulardan aynı sonuçlara varmayacağını varsaymalısınız. Dahası karar alıcı olarak, durumun ya da karar alma sürecinin alakasız yönleri olsa, belki de farklı bir karar alabileceğinizin farkında olmalısınız.

Başka bir deyişle tekil bir kararda gürültüyü ölçemeyiz ama karşıolgusal düşünürsek, gürültünün varlığından kesin olarak emin oluruz. Nişancının titrek elinin tek bir atışın başka bir yere de isabet *edebileceğini* düşündürmesi gibi, karar alıcılarda ve karar alma sürecindeki gürültü tekil kararın farklı *olabileceğini* düşündürür.

Tekil bir kararı etkileyen bütün faktörleri düşünün. Ebola tehdidini analiz etmekle ve tepki planları hazırlamakla sorumlu kişiler farklı köken ve eğitimlere sahip, farklı hayat tecrübelerinden gelen insanlar olsalardı, Başkan Obama'ya sundukları öneri aynı olur muydu? Aynı olgular sadece biraz farklı bir şekilde sunulsaydı, diyalog aynı şekilde mi ilerlerdi? Kilit oyuncular farklı bir ruh hâlinde olsalardı ya da bir kar fırtınası sırasında toplansalardı, nihai karar aynı olur muydu? Bu ışıkta baktığımızda, tekil karar o

kadar kati görünmez. Farkında bile olmadığımız pek çok faktöre bağlı olarak, karar, akla yatkın bir şekilde, farklı olabilirdi.

Yine bir karşıolgusal düşünme egzersizi olarak farklı ülke ve bölgelerin COVID-19 krizine nasıl farklı şekillerde tepki verdiğini düşünün. Virüsün onları kabaca aynı zaman ve benzer şekillerde vurmasına rağmen, tepkiler çok çeşitliydi. Bu çeşitlilik farklı ülkelerin muhakemelerindeki gürültünün açık bir kanıtıdır. Peki, ya salgın tek bir ülkeyi vursaydı? O zaman herhangi bir çeşitlilik gözlemlemeyecektik. Ancak bizim çeşitlilik gözlemleyememiz, kararın gürültüsünü azaltmayacaktı.

Tekil Kararlarda Gürültüyü Kontrol Etmek

Bu kuramsal tartışma çok önemli. Tekil kararlar tekrarlı kararlar kadar gürültülüyse, o zaman tekrarlı kararlarda gürültüyü azaltan stratejiler tekil kararların kalitesini de iyileştirmelidir.

Bu, mantığa göründüğünden daha aykırı bir reçete. Türünün tek örneği bir karar almanız gerektiğinde, içgüdünüz de muhtemelen ona türünün tek örneği gibi davranacaktır. Hatta kimileri, olasılığa dayalı düşünmenin kurallarının belirsizlik altında alınan tekil kararlarla tamamen alakasız olduğunu ve bu tür kararların kökten farklı bir yaklaşım gerektirdiğini iddia ederler.

Buradaki gözlemlerimiz, tam tersi bir öneri sunuyor. Gürültüyü azaltma açısından *tekil bir karar, sadece bir kez gerçekleşen tekrarlı bir karardır.* Bir kararı ister bir kez ister yüz kez alın, amacınız bunu hem ön yargıyı hem gürültüyü azaltacak şekilde yapmak olmalıdır. Dahası hatayı azaltan pratikler, türünün tek örneği kararlarınızda da tekrarlı kararlarınızdaki kadar etkili olmalıdır.

Tekil Kararlardan Bahsederken

"Bu sıra dışı fırsata yaklaşım şekliniz sizi gürültüye maruz bırakır."

"Unutmayın: Tekil bir karar, sadece bir kez gerçekleşen tekrarlı bir karardır."

"Sizi olduğunuz kişi yapan şahsi tecrübeler bu kararla gerçek anlamda alakalı değildir."

KISIM 2

Zihniniz Bir Ölçü Aletidir

Ölçüm, bilimde olduğu gibi gerçek hayatta da bir nesneye ya da olaya, bir alet kullanarak, bir ölçek üstünde değer tanımlama eylemidir. Bir halının uzunluğunu bir metre kullanarak santimle ölçersiniz. Sıcaklığı bir termometreden yardım alarak Fahrenheit ya da Celcius derece olarak ölçersiniz.

Bir yargıya varma eylemi de buna benzer. Yargıçlar bir suç için uygun hapis cezasını belirlerken, bir ölçek üstünde bir değer tanımlarlar. Risk değerlendirme uzmanlarının bir riski sigortalamak için dolar olarak değer belirlerken ya da doktorların teşhis koyarken yaptığı da budur. (Ölçek illa sayısal bir değer olacak diye bir kaide yoktur; "şüpheye yer bırakmayacak kadar suçlu", "ileri melanom" ve "cerrahi müdahale önerilir" de, bu tür bir ölçek anlatan yargılardır.)

Bu nedenle yargı *aletin insan zihni olduğu bir ölçüm* olarak tarif edilebilir. Ölçüm kavramında doğruluk hedefi kastedilir: gerçeğe yaklaşmak ve hatayı asgari düzeye çekmek. Yargının amacı etkilemek, bir tavır takınmak, ikna etmek değildir. Burada kul-

landığımız hâliyle yargı kavramının teknik psikolojik literatürden ödünç alındığına ve gündelik dildeki anlamından çok daha dar bir kavram olduğuna dikkatinizi çekmek isteriz. *Yargı, düşünmenin* eş anlamlısı olmadığı gibi *doğru yargılara varmak* da *iyi bir muhakeme yeteneğine sahip olmak*la eş anlamlı değildir.

Bizim tanımladığımız hâliyle yargı, bir sözcük ya da cümleyle özetlenebilecek bir çıkarımdır. Bir istihbarat analisti bir rejimin istikrarsız olduğu sonucuna ulaşan uzun bir rapor yazdığında, sadece sonuç yani çıkarım bir yargıdır. *Yargı, ölçüm* gibi hem zihin yoluyla bir yargıya varma aktivitesine hem de ürününe gönderme yapar. Bu anlamda *yargıç* kelimesini, bazen hukukla hiçbir ilgileri olmasa da yargıya varan insanlar anlamında da kullanacağız.

Her ne kadar amaç doğruluk olsa da, bu amaca ulaşmada bırakın yargıyı, bilimsel ölçümde bile mükemmellik sağlanamadı. Her zaman bir kısmı ön yargıdan, bir kısmı gürültüden kaynaklanan bir miktar hata olur.

Gürültü ve ön yargının hataya nasıl katkı sağladığını tecrübe etmeniz için sizi sadece bir dakikanızı alacak bir oyun oynamaya davet ediyoruz. Kronometresi olan bir akıllı telefonunuz varsa muhtemelen ardışık zaman aralıklarını kronometreyi durdurmadan, hatta ekranına bile bakmadan ölçmenizi sağlayan tur hafızası da vardır. Amacınız telefona bakmadan üst üste tam onar saniyelik beş ardışık tur üretmek. Başlamadan önce on saniyelik zaman dilimini birkaç kez gözlemlemek isteyebilirsiniz. Başlayın.

Şimdi telefonunuza kaydedilen tur sürelerine bakın. (Telefonun kendisi de gürültüden tamamen arınmış değildi ama gürültünün miktarı çok azdı.) Turların her birinin tam on saniye olmadığını ve hatırı sayılır bir yelpazede farklılık sergilediklerini göreceksiniz. Bire bir zamanlamayı üretmeye çalıştınız ama bunu yapamadınız. Kontrol edemediğiniz değişken, bir gürültü örneğidir.

Bu bulgu şaşırtıcı sayılmaz çünkü gürültü, fizyoloji ve psikolojide evrenseldir. Bireyler arasındaki farklılık, biyolojik bir sabittir;

bir elmanın iki yarısı bile gerçekten aynı değildir. Aynı insanın içinde de farklılıklar vardır. Nabzınız düzenli değildir. Aynı jesti kusursuz bir kesinlikle tekrarlayamazsınız. Bir odyoloji uzmanı sizi muayene ettiğinde, asla duymadığınız çok yumuşak sesler ve her zaman duyduğunuz çok yüksek sesler olacaktır. Öte yandan bazen duyacağınız, bazen kaçıracağınız birtakım sesler de olacaktır.

Şimdi telefonunuzdaki beş süreye bakın. Bir örüntü görebiliyor musunuz? Örneğin beş turun beşi de iç saatinizin hızlı çalıştığını düşündürecek şekilde, on saniyeden kısa mı? Bu basit görevde ön yargı, olumlu ya da olumsuz, turlarınızın ortalaması ve on saniye arasındaki farktır. Gürültü, daha önce gördüğümüz atışların dağılımına benzer şekilde, sonuçlarınızın değişkenliğini oluşturur. İstatistikte değişkenliğin en yaygın ölçüsü *standart sapma*dır ve biz de yargılardaki gürültüyü ölçmek için onu kullanacağız.

Çoğu yargıyı, özellikle ö*ngörücü* yargıları az önce yaptığınız ölçüme benzer bir şekilde düşünebiliriz. Bir öngörüde bulunduğumuz zaman, gerçek bir değere yaklaşmaya çalışırız. Bir ekonomi tahmincisi, bir sonraki yılın gayrisafi millî hasılasındaki büyümenin gerçek değerine olabildiğince yaklaşmayı hedef alır; bir doktor doğru teşhis koymayı amaçlar. (*Öngörü* teriminin bu kitapta kullanılan teknik anlamıyla geleceğe dair tahminde bulunmak anlamına gelmediğini belirtelim. Bizim amaçlarımız açısından, var olan bir hastalığın teşhisi de bir öngörüdür.)

Yargı ve ölçüm arasındaki benzerlikten çok yardım alacağız çünkü bu benzerlik, hatada gürültünün rolünü açıklamayı kolaylaştırır. Öngörücü yargılarda bulunan insanlar, hedef tahtasında boğa gözüne nişan alan nişancılara ya da bir partikülün gerçek ağırlığını ölçmeye çabalayan bir fizikçiye benzer. Yargılarındaki gürültü hatayı işaret eder. Basit bir ifadeyle, bir yargı gerçek bir değeri hedeflediğinde, iki farklı yargı doğru olamaz. Ölçüm aletleri gibi bazı insanlar belli görevlerde, belki beceri ya da eğitim eksiklikleri yüzünden, diğer insanlardan daha fazla hata sergiler. Ancak

yine ölçüm aletlerinde olduğu gibi, yargıda bulunan insanlar da hiçbir zaman kusursuz değildir. Hatalarını anlamalı ve ölçmeliyiz.

Elbette çoğu profesyonel yargı, bir zaman aralığını ölçmekten çok daha karmaşıktır. Bölüm 4'te farklı profesyonel yargı tiplerini tanımlayacak ve neyi hedeflediklerini inceleyeceğiz. Bölüm 5'te hatayı ve sistem gürültüsünün hataya katkısını nasıl ölçeceğimizi ele alacağız. Bölüm 6, sistem gürültüsünün daha derinine inerek farkı gürültü türleri olan ögelerini tanımlayacak. Bölüm 7'de bu ögelerden birini, durum gürültüsünü inceleyeceğiz. Son olarak Bölüm 8'de grupların yargıda gürültüyü genellikle nasıl katlayarak büyüttüğünü göstereceğiz.

Bu bölümlerden basit bir sonuç çıkıyor: Bir ölçüm aleti gibi insan zihni de kusurludur; hem yanlı hem de gürültülüdür. Neden ve ne kadar? Şimdi bunu görelim.

BÖLÜM 4

Yargı Meseleleri

Bu kitabın konusu genel anlamda profesyonel yargılardır ve bu tür bir yargıda bulunan herkesin yetkin olduğunu ve bunu doğru bir şekilde yapmayı hedeflediğini varsayar. Ancak yargı kavramının kendisi, bir yargının doğru olduğundan asla emin olunamayacağına dair gönülsüz bir onay içerir.

"Yargı meselesi" ya da "kişisel yargı" ifadelerini bir düşünün. Yarın güneşin doğacağını ya da sodyum klorürün formülünün NaCl olduğu önermelerini bir yargı meselesi olarak görmeyiz çünkü makul insanların bunlara tamamen katılması beklenir. Yargı meselesi denen şey, cevap konusunda bir belirsizlik olduğu ve makul ve becerikli insanların itiraz etmesi olasılığına izin verdiğimiz durumlardır.

Ancak kabul edilebilir itirazın bir sınırı vardır. Gerçekten de *yargı* kelimesi daha çok, insanların aynı görüşte olmaları gerektiğine inandıkları yerde kullanılır. Yargı meseleleri, çözümlenmemiş farklılıkların tamamen kabul edilebilir olduğu görüş ya da zevk meselelerinden farklıdır. Eğer sigorta eksperleri Beatles ve Rolling Stones'un meziyetleri ya da somon ve ton balığı konusunda keskin

bir şekilde ikiye ayrılmış olsaydı, gürültü denetiminin sonucuyla şoke olan sigorta yöneticilerinin bir sıkıntısı olmazdı.

Profesyonel yargılar dâhil yargı meseleleri, bir tarafta olgular ve hesaplama, diğer tarafta zevk ve görüş meseleleri arasında bir yer kaplar. *Sınırlı anlaşmazlık beklentisiyle* tanımlanırlar.

Bir yargıda tam olarak ne kadar anlaşmazlığın kabul edilebilir olduğu da başlı başına kanaate dayalı bir yargıdır ve sorunun zorluk derecesine bağlıdır. Bir yargı saçma olduğu zaman, bu yargı üzerinde bir konsensüs sağlamak özellikle kolaydır. Sıradan bir dolandırıcılık davasında verdikleri cezalarda büyük farklılıklar sergileyen yargıçlar, bir dolarlık bir para cezasının ve müebbet hapsin mantıksız olduğu konusunda görüş birliğine varacaklardır. Şarap yarışmalarında jüriler, hangi şarapların madalya alması gerektiği konusunda görüş farklılıkları sergiler ancak reddedilen şaraplarla ilgili olumsuz görüşleri genellikle oy birliğiyle ifade bulur.

Yargı Tecrübesi: Bir Örnek

Yargı tecrübesi konusunda ilerlemeden önce sizden kendinize bir tecrübe yaratmanızı istiyoruz. Bu egzersizi yapar ve sonuna kadar götürürseniz, bu bölümün geri kalanını daha iyi özümseyebilirsiniz.

> *Gittikçe artan bir rekabetle karşı karşıya olan vasat düzeyde başarılı bölgesel bir finans şirketinde CEO pozisyonu için adayları değerlendirmekle görevli bir ekipte olduğunuzu düşünün. Sizden aşağıdaki adayın iki yıl sonra görevde başarılı olma olasılığını değerlendirmeniz isteniyor. Başarı, basitçe adayın iki yılın sonunda CEO olarak pozisyonunu koruması şeklinde tanımlanıyor. Olasılığı 0 (imkânsız) ile 100 (kesin) bir ölçekte ifade edin.*
>
> *Michael Gambardi otuz yedi yaşında. On iki yıl önce Harvard İşletme Fakültesi'nden mezun olduğun-*

dan beri birçok görevde çalıştı. İlk yıllarda çok fazla finansal destek alamayarak başarısız olan iki start-up şirketi kurucusu ve yatırımcısıydı. Daha sonra büyük bir sigorta şirketine girdi ve kısa süre içinde şirketin Avrupa bölgesel operasyon direktörlüğüne yükseldi. Bu pozisyonda şikâyetlerin zamanında çözülmesi konusunda önemli bir ilerleme başlattı ve yönetti. Çalışma arkadaşları ve astları tarafından etkin olduğu kadar otoriter ve yıpratıcı olarak da tarif ediliyordu; görev süresi boyunca üst düzey yöneticilerde çarpıcı bir devir yaşandı. Çalışma arkadaşları ve astları ayrıca dürüstlüğüne ve başarısızlıklar için sorumluluk almaya istekli olduğuna da şahitlik ediyorlardı. Son iki yıl boyunca başlangıçta batma riski olan orta ölçekli bir finans şirketinde CEO olarak görev yaptı. Yeniden istikrar sağladığı şirkette başarılı ancak birlikte çalışması güç biri olarak tanımlanıyordu. Birkaç sene önce onunla mülakat yapan insan kaynakları uzmanları, yaratıcılık ve enerji konusunda yüksek puanlar verirken onu kibirli ve bazen zorba olarak tarif etmişlerdi.

Michael'in vasat ve gittikçe artan bir rekabetle karşı karşıya olan bölgesel bir finans şirketinde CEO adayı olduğunu hatırlayın. Michael'in, eğer işe alınırsa, iki yıl sonra hâlâ bu görevde olma olasılığı nedir? Lütfen okumaya devam etmeden önce, kararınızı 0 ile 100 arasında bir sayıyla belirtin. İhtiyaç duyarsanız, tarifi bir kez daha okuyun.

Bu göreve kendinizi ciddiyetle verdiyseniz, muhtemelen zor bulmuşsunuzdur. Büyük kısmı zahiren tutarsız olan çok fazla bilgi var. Bir yargı üretmek için ihtiyacınız olan tutarlı izlenimi

oluşturmakta hayli zorlandınız. O izlenimi oluştururken önemli görünen bazı detaylara odaklandınız ve büyük olasılıkla diğerlerini yok saydınız. Sizden seçtiğiniz sayıyı açıklamanız istense, birkaç çarpıcı olgudan bahsedersiniz ama yargınızı tam olarak açıklamaya yetecek kadarını saymazsınız.

İçinden geçtiğiniz düşünce süreci, yargı olarak adlandırdığımız zihinsel operasyonun birçok özelliğini resmeder.

Tarifin sağladığı bütün ipuçları arasında (ki bilmeniz gerekenlerin sadece bir alt kümesiydiler) bazılarına, yaptığınız seçimlerin tam anlamıyla farkında olmadan diğerlerinden daha fazla dikkat ettiniz. Gambardi'nin bir İtalyan adı olduğuna dikkat ettiniz mi? Okuduğu okulu hatırlıyor musunuz? Bu egzersiz, vakanın bütün detaylarını kolayca hatırlayamamanız için sizi aşırı yükleyecek şekilde tasarlandı. Muhtemelen sunduklarımızdan sizin hatırınızda kalanlarla başka okuyucuların aklında kalanlar birbirinden farklı olacaktır. Seçici dikkat ve seçici hafıza insanlar arasında bir farklılık kaynağıdır.

Ardından bu ipuçlarını *herhangi bir kural gözetmeksizin* Gambardi'nin beklentilerine ilişkin genel bir izlenimle birleştirdiniz. Burada anahtar kelime *herhangi bir kural gözetmeksizin*. Soruyu cevaplamak için bir plan oluşturdunuz. Yaptığınızın tam olarak farkında olmadan, zihniniz Michael'in güçlü yanları ve zayıf yanları ve karşı karşıya kaldığı zorluklar hakkında tutarlı bir izlenim yaratmak için çalıştı. Bu herhangi bir kural gözetmeme hızlı çalışmanıza olanak sağladı. Aynı zamanda çeşitlilik de yaratır: Bir sütun sayı eklemek gibi kurallı bir süreç aynı sonuçları garantilerken, herhangi bir kuralın gözetilmediği operasyonlarda bir miktar gürültü kaçınılmazdır.

Son olarak genel izleniminizi bir başarı olasılık ölçeği üstünde bir sayıya dönüştürdünüz. Bir izlenimi 0 ile 100 arasındaki bir

sayıyla eşleştirmek, Bölüm 14'te tekrar döneceğimiz önemli bir süreçtir. Bir kez daha, tam olarak neden bu şekilde tepki verdiğinizi bilmiyorsunuz. Neden 61 ya da 69 yerine 65'i seçtiniz? Büyük olasılıkla bir noktada aklınıza bir sayı geldi. O sayının doğru gelip gelmediğini kontrol ettiniz ve gelmediyse, başka bir sayı aklınıza girdi. Sürecin bir kısmı yine insanlar arasında bir çeşitlilik kaynağıdır.

Karmaşık bir yargı sürecinde bu üç adımın her biri bir değişkenliğe neden olduğu için Michael Gambardi'yle ilgili cevaplarda çok miktarda gürültü bulmamıza şaşırmamalıyız. Birkaç arkadaşınızdan bu örneği okumalarını rica ederseniz, muhtemelen, Gambardi'nin başarı olasılığıyla ilgili tahminlerinizin, büyük olasılıkla geniş bir alana yayıldığını görürsünüz. Bu örneği 115 MBA öğrencisine gösterdiğimiz zaman, Gambardi'nin başarı olasılığıyla ilgili tahminleri 10 ile 95 arasında değişkenlik gösterdi. Bu, çok büyük miktarda gürültü demek.

Bu arada, kronometre egzersizinin ve Gambardi sorununun iki tür gürültüyü resmettiğini fark etmiş olabilirsiniz. Kronometreyle peş peşe denemelerde varılan yargıların değişkenliği tek bir yargıcın (sizin) için(iz)deki gürültüdür. Oysa Gambardi örneğindeki yargıların değişkenliği farklı yargıçlar arasındaki gürültüdür. Ölçüm terimi olarak, ilk sorun *kişinin kendi içindeki* güvenilirlik sorununu, ikincisi *kişilerin arasındaki* güvenilirlik sorununu resmeder.

Yargının Ulaşmayı Hedeflediği Şey: İç Sinyal

Gambardi sorusuna cevabınız, terimin tanımladığımız anlamıyla öngörücü bir yargıdır. Bununla birlikte yarın Bangkok'taki en yüksek hava sıcaklığının kaç olacağı, bu akşamki futbol maçının sonucu ya da bir sonraki başkanlık seçimini kimin kazanacağı

gibi yine öngörücü olarak adlandırdığımız diğer yargılardan çok önemli açılardan farklılık gösterir. Bu sorunlar hakkında bir arkadaşınızla görüş ayrılığına düşerseniz, bir noktada kimin haklı çıktığını öğrenirsiniz. Ama Gambardi konusunda görüş ayrılığına düşerseniz, zaman size kimin haklı olduğunu *gösteremez*. Sebebi çok basit: Gambardi diye biri yok.

Soru gerçek bir insanı konu alsaydı ve sonucu bilseydik bile, tekil olasılık yargısı (0 ve %100 dışında) onaylanamaz ve çürütülemez. Sonuç beklenen olasılığın ne olduğunu ortaya sermez. %90'lık bir olasılık tanınan bir olay gerçekleşmediğinde, olasılık yargısının illa kötü olduğu söylenemez. Sonuçta gerçekleşme ihtimali sadece %10 olan sonuçlar, zamanın %10'unda gerçekleşirler. Gambardi örneği iki ayrı sebepten *doğrulanamaz* öngörücü yargının örneğidir: Gambardi bir kurgu karakterdir ve cevap olasılığa dayalıdır.

Pek çok profesyonel yargı doğrulanamaz türdendir. Örneğin, çok büyük hatalar dışında, risk değerlendirme uzmanları belli bir poliçenin aşırı mı yoksa gereğinden çok düşük mü fiyatlandığını asla bilmezler. Başka öngörüler de koşullu oldukları için doğrulanamayabilirler. "Savaşa girersek, parçalanırız." önemli bir öngörüdür ama sınanmaması muhtemeldir (umarız). Ya da öngörüler, öngörüde bulunan profesyonellerden hesap sorulamayacak kadar uzun vadeli olabilir. Örneğin yirmi birinci yüzyıl sonunda ortalama hava sıcaklığı tahmini gibi.

Gambardi görevinin doğrulanamaz doğası ona yaklaşımınızı değiştirdi mi? Örneğin, kendinize Gambardi'nin gerçek mi kurgu mu olduğunu sordunuz mu? Sonucun metnin devamında açıklanıp açıklanmayacağını merak ettiniz mi? Öyle bile olsa bu bilginin açıklanmasının karşı karşıya bırakıldığınız sorunun cevabını vermeyeceğini düşündünüz mü? Muhtemelen hayır, çünkü siz soruyu cevaplarken bu değerlendirmeler alakalı görünmemişti.

Doğrulanabilirlik, yargı tecrübesini değiştirmez. Bir dereceye kadar, ifşa edilme korkusu zihni yoğunlaştırdığı için, çözümü yakında

açıklanacak bir sorun üzerinde daha fazla kafa yorabilirsiniz. Tam tersi saçmalık derecesinde varsayımsal bir soruna kafa yormayı da reddedebilirsiniz ("Gambardi'nin üç bacağı olsaydı ve uçabilseydi, daha iyi bir CEO olur muydu?"). Ancak genel olarak akla yatkın, varsayımsal bir sorunu, gerçek bir sorunu ele alacağınız şekilde değerlendirirsiniz. Bu benzerlik büyük kısmı uydurulmuş sorunlar kullanan psikolojik araştırmalar için önemlidir.

Bir sonuç olmadığı için –dahası muhtemelen kendinize bir sonuç olup olmayacağını sormadınız bile– o sonuca göre hatayı asgari düzeye çekmeye çabalamadınız. Doğru yargıya varmaya, cevabınızı ilan etme güvenine sahip olabileceğiniz bir sayı saptamaya çalıştınız. Elbette cevabınızdan dört kere altının yirmi dört olduğundan emin olduğunuz gibi yüzde yüz emin değildiniz. Bir miktar belirsizliğin farkındaydınız (ve göreceğimiz gibi, muhtemelen sizin tahmin ettiğinizden daha fazla belirsizlik vardı). Ancak bir noktada, artık ilerleme kaydetmediğinize karar verdiniz ve bir cevap belirlediniz.

Size doğru ya da en azından cevabınız olabilecek kadar doğru bir yargıya vardığınızı hissettiren neydi? Biz bu duygunun dış bilgilerle bir ilgisi olmayan, *yargının tamamlandığına dair içsel sinyal* olduğunu düşünüyoruz. Cevabınız kanıta yeterince uyar gibi göründüyse doğru gelmiştir. Bir 0 ya da 100 cevabı size o uyma hissini vermezdi. İma ettiği güven, size sağlanan karmaşık, muğlak ve çelişkili kanıtla tutarsızdı. Ama karar kıldığınız sayı her neyse, size ihtiyaç duyduğunuz tutarlılık hissini verdi. Yargının amacı, sizin de tecrübe ettiğiniz gibi, tutarlı bir çözüme ulaşmaktı.

Bu içsel sinyalin esas özelliği, tutarlılık hissinin yargı tecrübesinin bir parçası olmasıdır. Gerçek bir sonuca bağlı değildir. Sonuç olarak içsel sinyal, doğrulanamaz yargılar için olduğu kadar gerçek, doğrulanabilir yargılar için de erişilebilirdir. Bu durum, Gambardi gibi bir kurgu karakter hakkında bir yargıya varma-

nın neden gerçek dünya hakkında yargıya varmaya yakın bir his verdiğini açıklar.

Yargı Nasıl Değerlendirilir: Sonuç ve Süreç

Doğrulanabilirlik yargı tecrübesini olduğu gibi değiştirmez. Ancak olaydan sonra değerlendirilmesini değiştirir.

Doğrulanabilir yargılar, objektif bir gözlemci tarafından basit bir hata ölçümüyle, başka bir deyişle yargı ve sonuç arasındaki farkla puanlanabilir. Bir hava durumu tahmincisi, bugünün en yüksek hava sıcaklığının 21 derece olduğunu söylediyse ve en yüksek sıcaklık 18 derece olduysa, tahminci artı üç derecelik bir hata yapmış demektir. Tahmin edileceği gibi bu yaklaşım, Gambardi problemi gibi gerçek bir sonucu olmayan, doğrulanamaz yargılarda işe yaramaz. Bu durumda bir yargıyı iyi yapan şeyin ne olduğuna nasıl karar vereceğiz?

Bu sorunun cevabı, yargıları değerlendirmenin ikinci bir yolunun olmasıdır. Bu yaklaşım hem doğrulanabilir hem doğrulanamaz yargılar için geçerlidir. Yargı *sürecini* değerlendirmeyi içerir. İyi ya da kötü yargılardan bahsederken sonuçtan (örneğin Gambardi vakasında ürettiğiniz sayıdan) ya da süreçten (o sayıya ulaşmak için ne yaptığınızdan) bahsediyor olabiliriz.

Yargı sürecinin değerlendirilmesine yönelik bir yaklaşım, o sürecin çok sayıda örneğe uyarlandığında nasıl bir performans sergilediğini gözlemlemektir. Örneğin yerel seçimlerde çok sayıda adaya kazanma olasılığını yakıştıran bir politika tahmincisini ele alalım. Tahmincimiz yüz adayın %70 olasılıkla kazanacağını belirtmiş olsun. Bu adaylardan yetmişi gerçekten seçilirse, tahmincinin olasılık ölçeğini kullanmadaki becerisi hakkında iyi bir işaret edinmiş oluruz. Hiçbir olasılık yargısı tek başına doğru ya da yanlış ilan edilemese de yargılar grup hâlinde doğrulanabilirdir. Benzer bir şekilde belli bir grubun lehine ya da aleyhine olaşa-

bilecek ön yargı, en iyi elle tutulur miktarda vakanın istatistiki sonuçlarının incelenmesiyle saptanabilir.

Yargı süreciyle ilgili sorulabilecek bir başka soru da bu sürecin mantık ya da olasılık teorisi prensiplerine uyup uymadığıdır. Bilişsel yargı ön yargıları üzerine çok sayıda araştırma bu yönde sonuç vermiştir.

Sonucundan çok yargı sürecine odaklanmak, kurgusal sorunlar hakkında yargılar ya da uzun vadeli öngörüler gibi doğrulanabilir olmayan yargıların kalitesini değerlendirmeyi mümkün kılar. Onları bilinen bir sonuçla karşılaştıramayabiliriz ama yine de hatalı olup olmadıklarını görebiliriz. Ve yargıları sadece değerlendirmektense *iyileştirme* sorununa döndüğümüzde, sürece de odaklanacağız. Bu kitapta ön yargıyı ve gürültüyü azaltmak için önerdiğimiz bütün prosedürler, benzer vakalar grubu üstünde hatayı asgari düzeye indirecek yargı sürecini benimsemeyi hedefler.

Bir yargıyı değerlendirmenin iki yolunu karşılaştırdık: yargıyı bir *sonuçla* karşılaştırarak ya da ona uzanan *sürecin* kalitesini değerlendirerek. Yargı doğrulanabilir olduğunda, onu değerlendirmenin iki yolunun tek bir vakada farklı çıkarımlara ulaşabileceğini dikkate alın. Mümkün olan en iyi alet ve teknikleri kullanan becerikli ve dikkatli bir tahminci üç aylık bir enflasyon öngörüsü yaparken doğru sayıyı sık sık kaçırır. Öte yandan tek bir çeyreklik dilimde dart atan bir şempanze bile bazen haklı çıkacaktır.

Karar alma konusunun uzmanları, bu gerilimi çözmek için çok net tavsiyelerde bulunurlar: Tek bir vakanın sonucuna değil, sürece odaklanın. Ancak biz gerçek hayatta standart uygulamanın bu olmadığının farkındayız. Profesyoneller genellikle yargılarının doğrulanabilir sonuçlara ne kadar yakın olduğuyla değerlendirilirler ve onlara yargılarında neyi amaçladıklarını sorarsanız, verecekleri cevap yaklaşmak olacaktır.

Özetle insanların doğrulanabilir yargılarda genellikle çabaladıklarını iddia ettikleri şey, sonuca uyan bir öngörüdür. Aslında

erişmeye çalıştıkları ise, doğrulanabilirlikten bağımsız olarak, vakanın olguları ve yargı arasındaki tutarlılığın sağladığı içsel tamamlanma sinyalidir. Oysa normatif bakış açısıyla ulaşmaya çalıştıkları şey, benzer vakalar grubu üstünde en iyi yargıyı üretecek yargı süreci olmalıdır.

Değerlendirmeci Yargılar

Bu bölümde şu ana dek öngörücü yargı görevlerine odaklandık; ele alacağımız yargıların çoğu da bu türden. Öte yandan Yargıç Frankel ve federal yargıçların hükümlerindeki gürültüyü ele alan Bölüm 1'de bir başka yargı türünü daha inceledik. Bir suçluya hüküm giydirmek öngörü değildir. Hükmü suçun ciddiyetiyle eşleştirmeyi amaç edinen bir *değerlendirmeci yargı*dır. Bir şarap yarışmasının jüri üyeleri ve restoran eleştirmenleri değerlendirmeci yargılarda bulunurlar. Deneme yazılarını puanlayan profesörler, buz pateni müsabakalarındaki jüri üyeleri ve araştırma projelerine burs veren komiteler değerlendirmeci yargılarda bulunurlar.

Değerlendirmeci yargının farklı bir türü, çoklu seçenekleri ve aralarındaki alışverişi içeren kararlarda yer alır. İşe alım için adaylar arasında seçim yapanları, stratejik seçenekler üstünde karar vermek zorunda olanları, hatta Afrika'daki bir salgına nasıl tepki vereceklerini seçen başkanları düşünün. Şüphesiz bütün bu kararlar, –bir adayın ilk senesinde nasıl performans sergileyeceği, borsanın belli bir stratejik hamleye nasıl tepki vereceği ya da kontrolsüz bırakıldığında salgının ne kadar hızlı yayılacağı gibi– girdi sağlayan öngörücü yargılara dayanır. Ancak nihai kararlar çeşitli seçeneklerin artıları ve eksileri arasında alışverişleri gerektirir ve bu alışverişler değerlendirmeci yargılarla çözüme kavuşturulur.

Öngörücü yargılar gibi değerlendirmeci yargılar da sınırlı anlaşmazlık beklentisine yol açar. Kendine saygısı olan hiçbir federal yargıç, "En çok hoşuma giden ceza bu ve meslektaşlarımın

aksini düşünmesi umurumda değil", demez. Birden fazla stratejik seçenek arasından seçim yapan karar alıcılar, aynı bilgiye ve aynı amaca sahip çalışma arkadaşlarının ve gözlemcilerin onlarla hemfikir olmasını ya da en azından aralarında çok büyük görüş ayrılıklarının olmamasını bekler. Değerlendirmeci yargılar kısmen bu tür yargılara varanların değer ve tercihlerine bağlıdır ama sadece zevk ve görüş meselesinden ibaret değildirler.

Bu nedenle öngörücü ve değerlendirmeci yargıların arasındaki sınır bulanıktır ve hüküm veren insanlar genellikle bunun farkında değildirler. Cezai hüküm belirleyen yargıçlar ya da öğrencilerin denemelerini puanlayan üniversite hocaları görevlerine kafa yorar ve "doğru" cevabı bulmak için çabalar. Yargıları ve yargılarının gerekçeleri konusunda güven geliştirirler. Profesyoneller, yargıları öngörücü olduğunda da ("Bu yeni ürün ne kadar satar?") değerlendirmeci olduğunda da ("Asistanım bu yıl nasıl bir performans sergiledi?") kendilerini haklı çıkarmak için aynı şeyleri hisseder, aynı şekilde davranır ve aynı şekilde konuşurlar.

Gürültünün Nesi Sorun?

Öngörücü yargılarda gürültünün gözlenmesi, daima bir sorun olduğuna işaret eder. İki doktor bir teşhis konusunda anlaşmazlığa düştüğünde ya da iki öngörücü bir sonraki üç aylık dönemin satışları konusunda aynı görüşte olmadığında, en azından biri yanılıyordur. Hata birinin daha az becerikli, dolayısıyla yanılmaya daha yatkın olmasıyla veya başka bir gürültü kaynağıyla alakalı olabilir. Sebep ne olursa olsun doğru yargıya varamamak, bu bireylerin teşhislerine ve öngörülerine bel bağlayan kişiler için çok ciddi sonuçlar doğurabilir.

Değerlendirmeci yargılarda gürültü, çok başka bir nedenden ötürü sorun yaratır. Yargıçların aralarında yer değiştirebileceğinin varsayıldığı ve kısmen rastgele atandığı her sistemde, aynı davayla

ilgili büyük anlaşmazlıklar hakkaniyet ve tutarlılık beklentilerini çiğneyebilir. Aynı sanığa verilen hükümlerde büyük farklılıklar varsa, Yargıç Frankel'in ilan ettiği "keyfî zulümler" alanındayız demektir. Hüküm vermede bireyselliğin değerine inanan ve bir soyguncunun cezası üstünde anlaşmazlığa düşen yargıçlar bile, bir yargıyı piyangoya dönüştüren bir anlaşmazlık düzeyinin sorunlu olduğunu kabul edeceklerdir. Benzer bir durum aynı denemeye çok farklı notlar, aynı restorana farklı güvenlik puanları, aynı buz patenciye farklı puanlar verildiğinde ya da depresyondaki bir insan sosyal güvenlik sisteminin engelli yardımlarından faydalanırken, aynı durumdaki bir başka kişi hiçbir şey alamadığında da (daha az çarpıcı görünse de) geçerlidir.

Haksızlığın sadece küçük bir endişe oluşturduğu durumlarda bile sistem gürültüsü başka bir soruna daha neden olur. Değerlendirmeci yargılardan etkilenen insanlar, söz konusu yargıların yansıttığı değerlerin yargı sahiplerinin değil, sistemin değerleri olmasını bekler. Kusurlu bir dizüstü bilgisayardan şikâyet eden bir müşteri parasının tamamını geri alırken bir başkası sadece özürle geçiştiriliyorsa veya bir şirkette beş yıldır çalışan biri terfi talebinde bulunup istediğini alıyor ama performansı aynı olan bir diğer çalışan kibarca geri çevriliyorsa ortada çok büyük bir sorun var demektir. Sistem gürültüsü tutarsızlıktır ve tutarsızlık, sistemin güvenilirliğine hasar verir.

İstenmeyen ama Ölçülebilir Olan

Gürültüyü ölçmek için tek ihtiyacımız, aynı sorun için birden fazla yargıdır. Gerçek bir değeri bilmemize gerek yoktur. Giriş bölümündeki atış poligonu hikâyesinin resmettiği gibi, hedef tahtasının arkasına baktığımız zaman hedef görünmez ama atışların ne şekilde dağıldığını görürüz. Nişancıların aynı hedefe nişan aldığını bildiğimiz anda, gürültüyü ölçebiliriz. Gürültü denetiminin

yaptığı budur. Bütün tahmincilerimizden önümüzdeki üç ayın satışları hakkında tahminde bulunmalarını istersek tahminlerinin dağılması gürültüdür.

Ön yargı ve gürültü arasındaki fark, yargıları iyileştirme amacı açısından önemlidir. Haklılıklarını doğrulayamazken yargıları iyileştirebileceğimizi iddia etmek paradoksal görünebilir ama iyileştirebiliriz; işe gürültüyü ölçmekle başlarsak. Yargının amacının sadece doğruluk mu yoksa değerler arasında daha karmaşık bir alışveriş mi olduğundan bağımsız olarak gürültü, istenmeyen ama sıklıkla ölçülebilir bir şeydir. Dahası Kısım 5'te ele alacağımız gibi gürültü ölçülebildiğinde, azaltılması genellikle mümkündür.

Profesyonel Yargıdan Bahsederken

"Bu bir yargı meselesidir. İnsanların kusursuz bir şekilde hemfikir olmasını bekleyemezsin."

"Evet, bu bir yargı meselesidir ama bazı yargılar normlara o kadar uzaktır ki ister istemez yanlıştır."

"Adaylar arasında yaptığın seçim ciddi bir yargının değil, zevkin ifadesiydi."

"Bir karar hem öngörücü hem değerlendirmeci yargılar gerektirir."

BÖLÜM 5

Hatayı Ölçmek

Sürekli bir ön yargının yüksek maliyetli hatalar üretebileceği açıktır. Bir tartı ağırlığınıza sürekli ekleme yapıyorsa, heyecanlı bir müdür rutin olarak projelerin sonuçta aldıkları sürenin yarısında tamamlanacağını öngörüyorsa ya da çekingen bir üst düzey yönetici gelecekteki satışlar konusunda her yıl gereksiz yere kötümser davranıyorsa, sonuç çok sayıda ciddi hata olacaktır.

Gürültünün maliyeti yüksek hatalar da üretebildiğini gördük. Şayet bir müdür çoğunlukla projelerin aldıkları sürenin yarısında tamamlanacağını öngörüyorsa ve arada bir gerçek sürelerinin iki katı zamanda biteceklerini tahmin ediyorsa, bu müdürün "genellikle" haklı çıktığını söylemek bize yardımcı olmaz. Farklı hatalar birbirlerini silmez, üst üste binerler.

Bu nedenle ön yargı ve gürültünün hataya nasıl ve ne kadar katlı sağladığı sorusu, önemlidir. Bu bölüm bu soruyu cevaplamayı amaçlıyor. Temel mesajı çok açık: Her tür profesyonel yargıda doğruluğun hedef olduğu her zaman, *ön yargı ve gürültü toplam hatanın hesaplanmasında aynı rolü üstlenir.* Bazı örneklerde katkının daha büyük kısmını ön yargı sağlarken, diğerlerinde bu

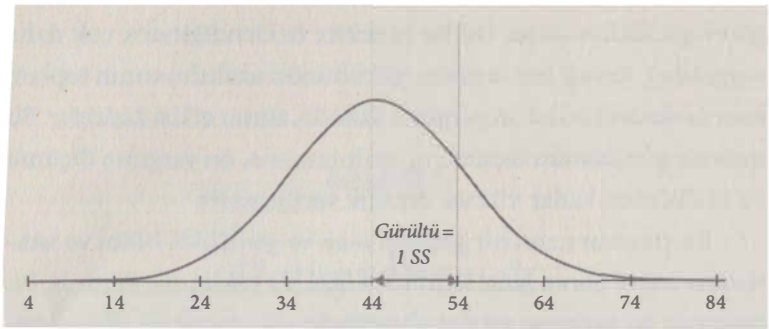
görev gürültüye düşer (ve bu örnekler beklendiğinden çok daha yaygındır). Ancak her örnekte, gürültünün azaltılmasının toplam hata üstündeki etkisi ön yargının azaltılmasının etkisi kadardır. Bu nedenle gürültünün ölçümü ve azaltılmasına, ön yargının ölçümü ve azaltılması kadar yüksek öncelik verilmelidir.

Bu çıkarım uzun bir geçmişi olan ve genellikle bilim ve istatistikte kabul gören hata ölçümüne özel bir yaklaşıma dayanır. Bu bölümde bu geçmişe giriş mahiyetinde göz atacak ve altta yatan akıl yürütme şeklinin bir eskizini çizeceğiz.

GoodSell Gürültüyü Azaltmalı mı?

GoodSell adında, çok sayıda satış tahmincisi çalıştıran büyük bir perakende satış şirketi hayal ederek başlayalım. Bu tahmincilerin görevi, GoodSell'in çeşitli bölgelerdeki piyasa payını öngörmek olsun. GoodSell'in tahmin departmanının başında bulunan Amy Simkin, belki de gürültü konusunda bir kitap okuduktan sonra, bir gürültü denetimi gerçekleştiriyor. Bütün tahminciler aynı bölgenin pazar payıyla ilgili bağımsız tahminler üretiyor.

Şekil 3 gürültü denetiminin (akıl almayacak kadar pürüzsüz) sonuçlarını gösteriyor. Amy tahminlerin hepimizin tanıdığı ve normal dağılım ya da Gauss dağılımı olarak da bilinen çan eğrisine dağıldığını görebiliyor. En sık tekrarlanan ve çan eğrisinin zirvesinin temsil ettiği tahmin %44. Amy şirketin tahmin sisteminin hayli gürültülü olduğunu görebiliyor. Hepsi doğru olsa aynı olması gereken tahminler çok geniş bir yelpazeye dağılmış durumda.



ŞEKİL 3: *GoodSell'in bir bölge için pazar payı tahminlerinin dağılımı*

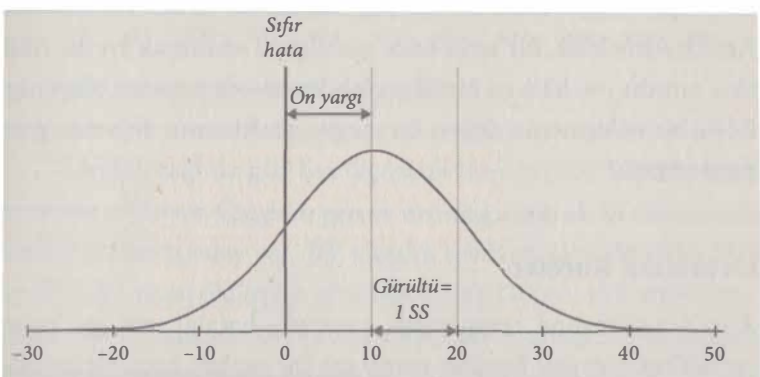
GoodSell'in tahmin sistemindeki gürültü miktarına bir sayı ekleyebiliriz. Turları ölçmek için kronometrenizi kullandığınız zaman yaptığımız gibi tahminlerin *standart sapmasını* hesaplayabiliriz. Adından da anlaşılacağı gibi standart sapma ortalamadan tipik bir uzaklığı temsil eder. Bu örnekte sapma %10 puan. Her normal dağılımda geçerli olduğu gibi tahminlerin üçte ikisi ortalamanın iki yanındaki birer standart sapma –bu örnekte %34 ve %54 pazar payı– içinde yer alıyor. Amy'nin elinde artık pazar payı tahminlerindeki sistem gürültüsünün miktarı konusunda bir tahmin var. (Daha iyi bir gürültü denetimi, daha sağlıklı bir tahmin için birkaç kestirim sorununu bir arada kullanırdı ama burada bizim amacımız için bir tane yeterli.)

Bölüm 2'deki gerçek sigorta şirketinin üst düzey yöneticilerinde olduğu gibi Amy de bu sonuçlar karşısında şoke oluyor ve eyleme geçmek istiyor. Kabul edilemez miktardaki gürültü tahmincilerin uymaları beklenen prosedürleri uygulamak konusunda disiplinli olmadıklarına işaret ediyor. Amy tahmincilerinin çalışmasında daha fazla birlik ve disipline ulaşmak için bir gürültü danışmanı tutmak için yetki istiyor. Ne yazık ki onay alamıyor. Patronunun cevabı yeterince mantıklı: Tahminlerimizin yanlış mı yoksa doğru mu olduğunu bilmeden hataları nasıl azaltabiliriz? Elbette, diyor, tahminlerde (büyük bir ön yargı gibi) büyük bir ortalama hata

varsa öncelik bunu ele almak olmalı. GoodSell'in tahminlerini iyileştirmek için herhangi bir girişimde bulunmadan önce, bekleyip doğru olup olmadıklarını görmesi gerektiğini belirtiyor.

İlk gürültü denetiminden bir yıl sonra, tahmincilerin öngörmeye çalıştığı sonuç artık biliniyor. GoodSell'in hedef bölgedeki pazar payı %34. Artık tahmincilerin her birinin hatasını da (sonuçla tahmin arasındaki fark) biliyoruz. %34'lük tahminde hata 0, ortalama tahmin olan %44'te hata %10 ve %24'lük düşük tahminde -%10.

Şekil 4 hataların dağılımını gösteriyor. Şekil 3'teki tahmin dağılımının aynısı ama gerçek değer (%34) her tahminden çıkarıldı. Dağılımın şekli değişmedi ve standart sapma (gürültü miktarı) hâlâ %10.



ŞEKİL 4: *GoodSell'in bir bölge için pazar payı tahminlerinde hata dağılımı*

Şekil 3 ve 4'ün arasındaki fark, Şekil 1 ve 2'de (bakınız Giriş Bölümü) atış deseninin hedef tahtasının ön ve arka yüzünden görünen hâli arasındaki farka benziyor. Atışlarda bir gürültü olduğunu gözlemlemek için hedefin konumunu bilmek gerekmiyordu. Benzer şekilde gerçek sonucu bilmek de tahmindeki gürültü konusunda zaten bilinene yeni bir şey eklemiyor.

Amy Silkin ve patronu daha önce bilmedikleri bir şeyi artık biliyorlar: tahminlerdeki ön yargı miktarı. Ön yargı en basit hâliyle bu örnekte %10 olan hata ortalamasıdır. Dolayısıyla ön yargı ve gürültü bu veri grubunda sayısal olarak aynı. (Daha net olması için, gürültü ve ön yargının eşitliği hiçbir şekilde genel bir kural değildir ancak ön yargı ve gürültünün eşit olduğu bir örnek rollerini anlamayı kolaylaştırır.) Çoğu tahmincinin iyimser bir hatta yaptığını, yani erişilecek pazar payını fazla tahmin ettiğini görüyoruz. Büyük kısmı sıfır hata dikey çizgisinin sağ tarafında hataya düştü. (Hatta normal dağılımın özelliklerini kullanarak, tahminlerinin %84'ünde durumun böyle olduğunu biliyoruz.)

Amy'nin patronu, pek gizlemediği bir hoşnutlukla ifade edeceği gibi, haklıydı. Tahminlerde çok fazla ön yargı vardı! Ve şimdi ön yargıyı azaltmanın iyi bir şey olacağı da netlik kazanmıştı. Ancak Amy hâlâ, bir sene önce gürültüyü azaltmak iyi bir fikir olur muydu –ve hâlâ iyi bir fikir olabilir mi– diye merak ediyordu. Böyle bir iyileşmenin değeri ön yargıyı azaltmanın değerine göre nasıl olurdu?

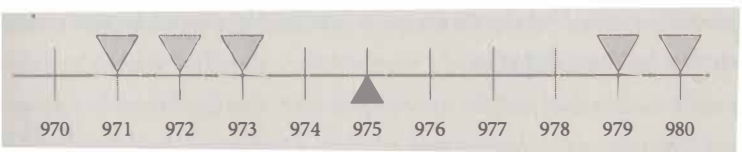
Ortalama Kareler

Amy'nin sorusunu cevaplamak amacıyla, hatalar için bir "puan cetveli"ne, bireysel hataları tartıp tek bir toplam hata ölçüsünde bir araya toplamaya ihtiyacımız var. Neyse ki böyle bir araç var: 1777 yılında dünyaya gelen ve çok büyük keşiflerle dolu bir kariyere henüz onlu yaşlarındayken adım atan ünlü matematik dehası Carl Friedrich Gauss tarafından 1795 yılında keşfedilen *en küçük kareler yöntemi*.

Gauss ayrı ayrı hataların, toplam hataya katkısını puanlamak için bir kural önerdi. Gauss'un toplam hata ölçümü –*ortalama karesel hata* (OKH) denir– bireysel ölçüm hatalarının karelerinin ortalamasıdır.

Gauss'un toplam hatanın ölçümüne yaklaşımı için detaylı argümanları, bu kitabın kapsamını fazlasıyla aşıyor ve çözümü ilk bakışta bariz gelmiyor. Hataların karelerini kullanmak neden? Fikir keyfî, hatta tuhaf görünüyor. Yine de göreceğiniz gibi neredeyse kesin olarak paylaştığınız bir önsezi üstüne kuruluyor.

Nedenini anlamak için tamamen farklı gibi görünen ama aslında aynısı olan bir soruna dönelim. Diyelim size bir cetvel verildi ve bir çizginin uzunluğunu en yakın milimetreyi gösterecek şekilde ölçmeniz istendi. Beş ölçüm yapma hakkınız var. Bu ölçümler Şekil 5'te çizginin üstündeki aşağı bakan üçgenlerle temsil ediliyor.



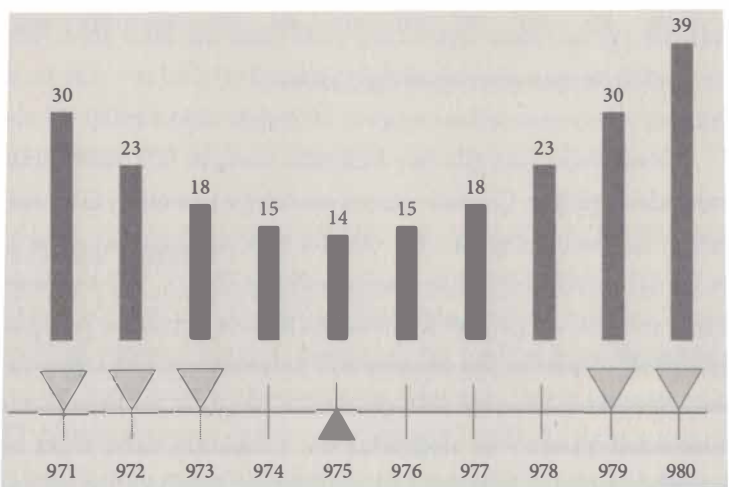
ŞEKİL 5: *Aynı uzunluğun beş ölçümü*

Görebileceğiniz gibi beş ölçümün tamamı 971 ve 980 metre arasında değişiyor. Çizginin gerçek uzunluğu için en iyi tahmininiz nedir? İki bariz aday var. Bir olasılık medyan sayı, iki daha kısa ve iki daha uzun ölçümün arasında duran ölçüm. 973 milimetre. Diğer olasılık ise günlük kullanımda sadece ortalama dediğimiz aritmetik ortalama. Bu örnekte 975 milimetreye denk geliyor ve yukarıyı işaret eden bir okla gösterildi. Sezginiz muhtemelen ortalamadan yanadır ve doğrudur da. Ortalama daha fazla bilgi içeriyor; sayıların boyutundan etkileniyor. Medyan ise sayıların sadece sıralanışlarından etkileniyor.

Hakkında net bir sezginizin olduğu bu tahmin sorunuyla burada üzerinde durduğumuz toplam hata sorunu arasında sıkı bir bağlantı var. İkisi aynı madenî paranın iki yüzü denebilir. Bunun nedeni, en iyi tahminin mevcut ölçümlerin toplam hatasını en aza indiren tahmin olmasıdır. Buna göre ortalamanın en

iyi tahmin olduğu konusunda sezginiz doğruysa, toplam hatayı ölçmek için kullandığınız formül aritmetik ortalamayı hatanın asgariye indirildiği değer olarak kullanan formül olmalı.

OKH bu özelliğe sahip ve toplam hatanın bu özelliğe sahip tek tanımı. Şekil 6'da beş ölçüm setinde, çizginin gerçek uzunluğunun on olası tam sayı değeri için OKH değerini hesapladık. Örneğin gerçek değer 971 olsaydı, beş ölçümdeki hatalar 0, 1, 2, 8 ve 9 olurdu. Bu hataların karelerinin toplamı 150 ve ortalamaları da 30 ediyor. Bu, bazı ölçümlerin gerçek değere çok uzak olduğunu yansıtan büyük bir sayı. 975'e (yani ortalamaya) yaklaştıkça OKH'nin azaldığını ve o noktanın ötesinde yeniden yükseldiğini görebilirsiniz. Ortalama en iyi tahminimiz çünkü toplam hatayı en aza indiren değer o.



ŞEKİL 6: *Doğru uzunluğun on olası değeri için ortalama karesel hata (OKH)*

Tahmininiz ortalamadan uzaklaştıkça toplam hatanın hızla arttığını da görebilirsiniz. Örneğin, tahmininiz 976'dan 979'a sadece 3 milimetre arttığında, OKH ikiye katlanır. Bu OKH'nin kilit

özelliklerinden biridir. Kare almak büyük hatalara küçük hatalara verdiğinden daha fazla ağırlık verir.

Gauss'un toplam hatayı ölçme formülüne neden ortalama karesel hata dendiğini ve tahmine yaklaşımına neden en küçük kareler yöntemi dendiğini artık anlamışsınızdır. Hataların karesinin alınması bu yöntemin ana fikridir ve başka hiçbir formül, ortalamanın en iyi tahmin olduğu sezginizle bağdaşmazdı.

Gauss'un yaklaşımının avantajları diğer matematikçiler tarafından kısa süre içinde anlaşıldı. Gauss, daha pek çok başarısının yanı sıra OKH'yi, Avrupa'nın en iyi astronomlarını alt eden bir bilmeceyi, daha önce 1801 yılında, güneşin parlak ışığında gözden kaybolmadan önce sadece kısa süre izi sürülebilen bir asteroit olan Ceres'in yeniden keşfini çözmekte kullandı. Astronomlar Ceres'in yörüngesini tahmin etmeye çalışıyorlardı ama teleskoplarının ölçüm hatasını açıklama şekilleri hatalıydı ve gezegen onların sonuçlarının öne sürdüğü konumun yakınlarında yeniden görünmedi. Gauss hesaplamaları en küçük kareler yöntemini kullanarak tekrarladı. Astronomlar teleskoplarını onun işaret ettiği yere çevirerek Ceres'i buldular.

Çok çeşitli disiplinlerden bilim insanları en küçük kareler yöntemini benimsemekte gecikmediler. İki yüzyıl sonra bile bu yöntem, hassas doğruluğa ulaşmanın hedef olduğu durumlarda hataları değerlendirmenin standart yolu olarak kullanılıyor. Hataları kareleriyle tartmak istatistikte merkezî rol oynuyor. Bütün bilimsel disiplinlerdeki uygulamaların büyük kısmında OKH hâkimdir. Birazdan göreceğimiz gibi bu yaklaşımın şaşırtıcı sonuçları da var.

Hata Denklemleri

Hatada ön yargı ve gürültünün rolü *hata denklemleri* olarak adlandıracağımız iki ifadede kolayca özetlenebilir. Bu denklemlerin ilki hatayı tek bir ölçümde artık aşina olduğunuz iki ögeye çözümler: ön yargı –ortalama hata– ve artık "gürültülü hata".

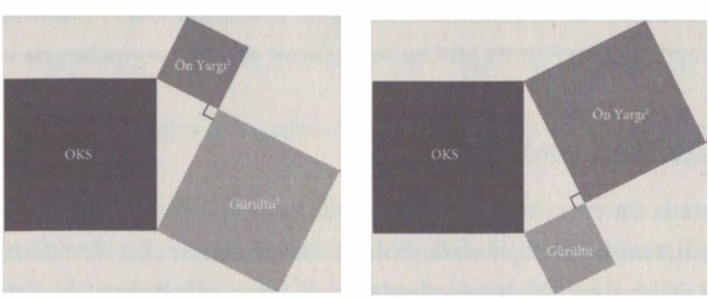
Gürültülü hata, hata ön yargıdan büyük olduğunda pozitif, küçük olduğunda ise negatiftir. Gürültülü hataların ortalaması sıfırdır. İlk hata denkleminde yeni bir şey yok.

Tekil Bir Ölçümde Hata = Ön Yargı + Gürültülü Hata

İkinci hata denklemi OKH'nin bir tür ayrıştırılması, şimdi devreye soktuğumuz toplam hatanın ölçümüdür. Basit cebir kullanılarak, OKH'nin ön yargı ve gürültünün karelerinin toplamına denk olduğu gösterilebilir. (Gürültünün ölçümlerin standart sapması olduğunu hatırlayalım.)

Toplam Hata (OKH) = Ön Yargı^2 + Gürültü^2

Bu denklemin biçimi –iki karenin toplamı– size lisede çok sevilen bir denklemi, Pisagor teoremini hatırlatabilir. Hatırlayacağınız gibi bir dik üçgende, iki kısa kenarın karelerinin toplamı, uzun kenarın karesine eşittir. Bu bize hata denkleminin, OKH, Ön Yargı^2 ve Gürültü^2'nin bir dik üçgenin kenarlarındaki üç karenin alanı olduğu basit bir görselleştirmesini sağlar. Şekil 7 OKH'nin (koyu renkli karenin alanı) diğer iki karenin alanlarının toplamına eşit olduğunu gösteriyor. Sol panelde ön yargıdan çok gürültü var, sağdakinde ise gürültüden çok ön yargı mevcut. Ancak OKH aynı ve hata denklemi her iki durumda da geçerliliğini koruyor.



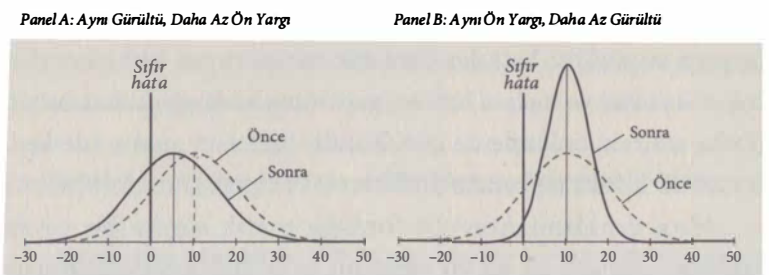
ŞEKİL 7: *OKH'nin iki çözümlemesi*

Hem matematiksel ifadenin hem görsel sunumun önerdiği gibi ön yargı ve gürültü, hata denkleminde eşit rol oynar. Birbirlerinden bağımsızdırlar ve toplam hatanın saptanmasında ağırlıkları eşittir. (Daha sonraki bölümlerde gürültünün ögelerini analiz ederken, benzer bir kareler toplamına çözümleme kullanacağımızı belirtelim.)

Hata denklemi Amy'nin sorduğu pratik soruya bir cevap sağlıyor: Gürültü ya da ön yargının aynı miktarda azaltılması toplam hatayı nasıl etkiler? Cevap çok açık: Ön yargı ve gürültü hata denkleminde yer değiştirebilir ve azalan hangisi olursa olsun, toplam hatadaki azalma aynı olacaktır. Ön yargı ve gürültünün eşit (her ikisinin %10) olduğu Şekil 4'te, toplam hataya katkıları eşittir.

Hata denklemi ayrıca, Amy Simkin'in başlangıçtaki gürültüyü azaltma dürtüsüne su götürmez bir destek sağlar. Gürültüyü gözlemlediğiniz zaman azaltmak için çabalamalısınız! Denklem Amy'nin patronunun Goodsell'in öngörülerindeki ön yargıyı ölçmek için beklemesini ve ne yapılacağına daha sonra karar vermelerini önerirken yanıldığını gösteriyor. Toplam gürültü açısından, gürültü ve ön yargı bağımsızdır; ön yargının miktarı ne olursa olsun, gürültüyü azaltmanın getirisi aynıdır.

Bu nosyon genel kanıya ters görünmekle birlikte çok önemlidir. Daha iyi anlamamız için, Şekil 8 ön yargı ve gürültüyü aynı miktarda azaltmanın etkisini gösteriyor. Her iki panelde elde edilenleri takdir etmenize yardımcı olmak için, hataların orijinal dağılımı (Şekil 4'ten) burada kesik çizgiyle gösterildi.



ŞEKİL 8: *Hataların ön yargının yarı yarıya azalması ve gürültünün yarı yarıya azalması durumunda dağılımı*

Panel A'da, Amy'nin patronunun dediğinin olduğunu varsayıyoruz: Ön yargının ne olduğunu buldu ve bir şekilde (belki aşırı iyimser öngörücülere geri besleme sağlayarak) yarıya indirmeyi başardı. Gürültü konusunda bir şey yapılmadı. Kaydedilen ilerleme görülebiliyor: Öngörülerin bütün dağılımı gerçek değere yaklaştı.

Panel B'de, tartışmayı Amy kazansa olacakları görüyoruz. Görüldüğü gibi ön yargı değişmedi belki ama gürültü yarı yarıya azaldı. Buradaki paradoks, gürültünün azalmasının durumu kötüleştirmiş gibi görünmesidir. Şimdi öngörüler daha konsantre (daha az gürültülü) ama daha doğru (daha az ön yargılı) değiller. Daha önce öngörülerin %84'ü gerçek değerin bir tarafındayken artık neredeyse tamamı (%98'i) gerçek değerin çok üstüne çıktı. Gürültünün azaltılması öngörüleri daha yanlış yapmış görünüyor: Amy'nin umduğu türde bir ilerlemenin kaydedildiği söylenemez!

Ancak görüntüye rağmen toplam hata, Panel B'de de Panel A'daki kadar azaldı. Panel B'deki kötüleşme yanılsaması ön yargıyla ilgili hatalı bir sezgiden doğuyor. İlgili ön yargı ölçümü olumlu ve olumsuz hataların dengesizliği değil. Çan eğrisinin zirvesiyle gerçek değer arasındaki mesafe olan ortalama hata. Panel B'de bu ortalama hesap orijinal duruma göre değişmedi; hâlâ yüksek, %10'da ama daha kötü değil. Doğru, ön yargının varlığı şimdi daha çarpıcı çünkü toplam hatanın daha büyük bir oranını açıklıyor

(%50 yerine % 80). Öte yandan bunun nedeni gürültünün azalmış olması. Tersine, panel A'da ön yargı azaldı ama gürültü azalmadı. Net sonuç OKH'nin iki panelde de aynı olması: Gürültüyü ya da ön yargıyı aynı miktarda azaltmak OKH üstünde aynı etkiyi yaratıyor.

Bu örneğin gösterdiği gibi OKH, öngörücü yargıların puanlanmasıyla ilgili yaygın sezgilerle çelişiyor. OKH'yi en aza indirmek için, büyük hatalardan kaçınmaya konsantre olmalısınız. Örneğin uzunluk ölçerken bir hatayı 11 cm'den 10 cm'ye düşürmenin etkisi, 1 cm'lik bir hatadan tam isabete ulaşmaktan yirmi kat daha büyüktür. Ne yazık ki insanların bu açıdan sezgileri neredeyse her zaman olması gerekenin ayna görüntüsüdür: İnsanlar tam isabete ulaşmak için çok hevesli ve küçük hatalara karşı fazlasıyla duyarlıdırlar ama iki büyük hata arasındaki farka pek aldırmazlar. Hedefinizin doğru yargılara ulaşmak olduğuna samimiyetle inanıyorsanız bile, sonuçlara duygusal tepkiniz bilimin tanımladığı hâliyle hatasızlığa ulaşmakla bağdaşmayabilir.

Elbette burada en iyi çözüm hem gürültüyü hem ön yargıyı azaltmak olurdu. Ön yargı ve gürültü bağımsız oldukları için, Amy Simkin ve patronu arasında seçim yapmak için bir neden yoktur. Bu açıdan eğer GoodSell gürültüyü azaltmaya karar verirse, gürültünün azaltılmasının ön yargıyı daha görünür –hatta gözden kaçırılmasını imkânsız– kılması bir armağan olabilir. Gürültüyü azaltmayı başarmak, ön yargıyı azaltmanın şirketin bir sonraki gündemi olmasını garantileyecektir.

Hiç kuşkusuz ön yargı gürültüden daha büyük olduğunda, gürültüyü azaltmak daha az öncelikli olurdu. Ancak GoodSell örneği vurgulamaya değer bir ders daha veriyor. Bu basitleştirilmiş modelde, gürültü ve ön yargının eşit olduğunu varsaydık. Hata denkleminin biçimi düşünüldüğünde, toplam hataya katkıları da eşittir: Ön yargı, toplam hatanın %50'sini açıklar, gürültü de öyle. Bununla birlikte, daha önce de belirttiğimiz gibi, öngörücülerin %84'ü aynı yönde hataya düşüyor. Gürültü kadar büyük bir etki

yaratması için bu kadar büyük bir ön yargı (yedi kişiden altısının aynı yönde hata yapması) gerek. Bu nedenle ön yargıdan çok gürültünün olduğu durumlar bulmak bizi şaşırtmamalı.

Burada hata denkleminin tek bir vakaya, GoodSell'in alanının tek bir bölgesine uygulanmasını gösterdik. Elbette her zaman aynı anda birden fazla örnekte gürültü denetimi yapmak tercih edilir. Değişen bir şey olmaz. Hata denklemi ayrı örneklere uygulanır, toplam bir denklem OKH, ön yargı kare ve gürültü karenin vakalar üstündeki ortalaması alınarak elde edilir. Amy Simkin'in birden fazla bölge için, aynı ya da farklı öngörücülerden çok sayıda öngörü alması daha iyi olurdu. Ortalama sonuçlar ona GoodSell'in öngörü sistemindeki ön yargı ve gürültünün daha doğru bir resmini sağlardı.

Gürültünün Maliyeti

Hata denklemi bu kitabın entelektüel temelini oluşturuyor. Öngörücü yargılarda sistem gürültüsünü azaltma hedefinin, prensipte istatistiksel ön yargının azaltılması kadar önemli olan bir hedefin gerekçesini sağlıyor. (İstatistiksel ön yargının sosyal ayrımcılığın eş anlamlısı olmadığını vurgulamalıyız. İstatistiksel ön yargı, sadece bir yargı grubundaki ortalama hatadır.)

Hata denklemi ve ondan çıkardığımız sonuçlar, OKH'nin toplam hata ölçüsü olarak kullanımına bağlıdır. Kural, her biri gerçek bir değere maksimum isabet (en az ön yargı) ve netlikle (en az gürültü) yaklaşmayı hedefleyen tahmin ve öngörüler dâhil tamamen öngörücü yargılar için uygundur.

Öte yandan hata denklemi, değerlendirmeci yargılara uymaz çünkü gerçek bir değerin varlığına bağlı olan hata kavramını uygulamak çok daha zordur. Dahası hatalar belirlenebilseydi bile, maliyetleri nadiren simetrik olurdu ve karelerine kesin olarak orantılı olma ihtimalleri pek olmazdı.

Örneğin asansör üreten bir firma için bir asansörün maksimum taşıma kapasitesini tahminde yapılan hataların sonuçları bariz bir şekilde asimetriktir: Gerçek değerin altında kalmak maliyetlidir ama üstüne çıkmak faciaya yol açabilir. Kare hata bir trene yetişmek için evden ne zaman çıkmak gerektiği kararıyla da, benzer şekilde alakasızdır. Bu kararda, bir dakika ya da beş dakika geç kalmanın sonuçları aynıdır. Bölüm 2'deki sigorta şirketi poliçeleri fiyatlandırırken ya da hasar taleplerinin değerini tahmin ederken, her iki yöndeki hatalar da maliyetlidir ama maliyetlerinin denk olacağını varsaymak için bir neden yoktur.

Bu örnekler kararlarda öngörücü ve değerlendirmeci yargıların rollerini belirleme ihtiyacını vurguluyor. İyi karar alma konusunda yaygın olarak kabul gören bir düstur, değerlerinizi olgularınızla karıştırmamanız gerektiğidir. İyi kararlar umutlardan ve korkulardan ya da tercihlerden ve değerlerden hiçbir şekilde etkilenmeyen nesnel ve doğru öngörücü yargılara dayanmalıdır. Asansör firması için ilk adım, asansörün maksimum teknik taşıma kapasitesinin farklı mühendislik çözümleri altında nötr olarak hesaplanması olurdu. Güvenlik sadece, değerlendirmeci yargının maksimum kapasiteyi saptamak için kabul edilebilir bir güvenlik payı tercihini belirlediği ikinci adımda baskın endişe olur. (Emin olmak adına, o seçim büyük ölçüde örneğin, o güvenlik payının maliyetleri ve getirilerini içeren olgusal yargılara bağlı olacaktır.) Benzer şekilde istasyona gitmek için evden ne zaman çıkacağınıza karar vermenin ilk adımı, farklı seyahat sürelerinin olasılıklarının objektif olarak saptanması olmalıdır. Treninizi kaçırmanın ve istasyonda zaman kaybetmenin ayrı ayrı maliyetleri sadece kabul etmeye razı olduğunuz risk seçiminizde önemli hâle gelir.

Aynı mantık sonucu çok daha önemli kararlarda da geçerlidir. Bir ordu komutanı bir saldırıyı başlatıp başlatmamaya karar verirken, birçok noktayı tartmalıdır ama liderin bel bağladığı istihbaratın büyük kısmı bir öngörücü yargı meselesidir. Salgın gibi bir sağlık

krizine tepki veren bir hükûmet çeşitli seçeneklerin artılarını ve eksilerini tartmalıdır ama her seçeneğin (hiçbir şey yapmama kararı dâhil) olası sonuçları hakkında doğru öngörüler olmadan, herhangi bir değerlendirme mümkün olmaz.

Bütün bu örneklerde, nihai kararlar değerlendirmeci yargıları gerektirir. Karar alıcılar birden çok seçeneği göz önünde bulundurmalı ve en uygun tercihi yapmak için değerlerini uygulamalıdırlar. Fakat kararlar altta yatan, değer açısından nötr olması gereken öngörülere dayanır. Öngörülerin amacı doğruluk –hedefin olabildiğince yakınına isabet etmek– ve OKH uygun hata ölçüsüdür. Öngörücü yargılar gürültüyü azaltan prosedürlerle iyileştirilir; tabii ön yargıyı büyük ölçüde artırmadıkları sürece.

Hata Denkleminden Bahsederken

"Tuhaftır ki ön yargıyı ve gürültüyü aynı miktarda azaltmak doğruluk üstünde aynı etkiyi yaratır."

"Öngörücü yargıda gürültüyü azaltmak, ön yargı hakkında ne bildiğinizden bağımsız olarak, her zaman faydalıdır."

"Yargılar gerçek değerin üstündeki ve altındakiler arasında 84 ve 16 olarak ayrıldığında, büyük bir ön yargı var demektir. Bu durumda ön yargı ve gürültü eşittir."

"Öngörücü yargılar her kararda bulunur ve doğruluk tek amaçları olmalıdır. Değerlerinizi ve olgularınızı ayrı tutun."

BÖLÜM 6

Gürültü Analizi

Bir önceki bölüm tek bir vakanın ölçümünde ya da yargısında değişkenliği ele almıştı. Tek bir vakaya odaklandığımız zaman, yargıdaki her değişkenlik hatadır ve hatanın iki ögesi ön yargı ve gürültüdür. Elbette, mahkemeleri ve sigorta şirketlerini içerenler dâhil incelediğimiz yargı sistemleri farklı vakalarla ilgilenmek ve aralarında ayrım yapmak üzere tasarlanmıştır. Federal yargıçlar ve sigorta eksperleri önlerine konan her vakada aynı yargıya varsalardı çok işe yaramazlardı. Farklı vakaların yargılarındaki değişkenliğin büyük kısmı kastidir.

Ancak aynı vakanın yargılarında değişkenlik arzulanan bir şey değildir, sistem gürültüsüdür. Göstereceğimiz gibi aynı insanların birden çok vaka hakkında yargıda bulunduğu bir gürültü denetimi, sistem gürültüsünün daha detaylı analizine olanak sağlar.

Hüküme Dair Bir Gürültü Denetimi

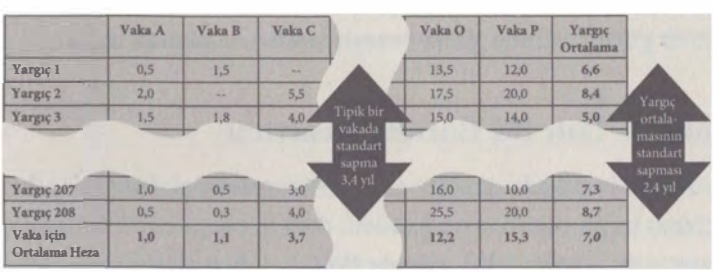
Birden fazla vakada gürültü analizini resmetmek için federal yargıçların hükümlerinin olağanüstü detaylı bir gürültü denetimine döneceğiz. Analiz 1981 yılında Bölüm 1'de anlattığımız hüküm reformuna doğru bir hareketin parçası olarak yayımlandı. Çalışma

dar bir alanda hüküm kararlarına odaklanmıştı ama sunduğu dersler genel mahiyette ve diğer profesyonel yargıları da ilgilendiriyor. Gürültü denetiminin amacı, Yargıç Frankel'in ve diğerlerinin bir araya topladığı canlı ama anekdota dayalı kanıtların ötesine geçmek ve "hükümlerde uyuşmazlığın boyutunu" daha sistematik olarak saptamaktı.

Çalışmanın yazarları sanığın suçlu bulunduğu ve hüküm giyeceği on altı farazi dava geliştirdiler. Özetler ya soygunları ya da dolandırıcılık vakalarını tasvir ediyor ve sanığın suçta fail mi yoksa suç ortağı mı olduğu, bir sabıka kaydının olup olmadığı ya da (soygun vakalarında) bir silahın kullanılıp kullanılmadığı gibi altı diğer boyutta farklılıklar sergiliyordu.

Araştırmacılar ülkenin farklı yerlerinden seçilmiş 208 aktif federal yargıçla özenle hazırlanmış mülakatlar organize ettiler. Doksan dakika içinde yargıçlara on altı vaka sunuldu ve bir hüküm belirlemeleri istendi.

Bu çalışmadan öğrenebilecekleri daha iyi anlamak için, bir görselleştirme egzersizi faydalı olacaktır. Zihninizde suçlar için A'dan P'ye kadar işaretlenmiş on altı sütunu ve yargıçlar için 1'den 208'e kadar numaralandırılmış 208 sırası olan bir tablo canlandırın. A1'de P208'e kadar her hücre belli bir vaka için, belli bir yargıç tarafından verilen hapis cezasını gösteriyor. Şekil 9 bu 3.328 hücreli tablonun neye benzeyeceğini resmediyor. Gürültüyü incelemek için her biri ayrı bir gürültü denetimi olan on altı sütuna odaklanacağız.



ŞEKİL 9: *Hüküm incelemesinin bir temsili*

Ortalama Hüküm Süreleri

Belli bir vaka için bir hükmün "gerçek değeri"ni belirlemenin objektif bir yolu yoktur. Aşağıda her vaka için 208 hükmün ortalamasını (ortalama hüküm süresini) o vaka için "doğru" hükümmüş gibi alacağız. Bölüm 1'de değindiğimiz gibi ABD Ceza Komisyonu hüküm verme kılavuzunun hazırlanmasında temel olarak geçmişteki ortalama uygulamayı alırken aynı varsayımda bulunmuştu. Bu nitelik, her vakanın ortalama hükmünde sıfır ön yargı varsayıyor.

Gerçekte bu varsayımın yanlış olduğunun bütünüyle farkındayız; bazı vakaların ortalama yargısının, fazlasıyla benzer başka vakaların ortalama yargısına göre –örneğin ırksal ayrımcılık nedeniyle– ön yargılı olması çok olasıdır. Vakalar arasındaki –kimi pozitif, kimi negatif– ön yargı çeşitliliği önemli bir hata ve haksızlık kaynağıdır. Bu değişkenlikten, kafa karıştırıcı bir şekilde sıklıkla "ön yargı" olarak bahsedilir. Bu bölümdeki –ve bu kitaptaki– analizimiz–ayrı bir hata kaynağı olan gürültüye odaklı. Yargıç Frankel gürültünün adaletsizliğini vurguladı ama aynı zamanda (ırk ayrımcılığı dâhil) ön yargıya da dikkat çekti. Benzer bir şekilde gürültüye odaklanmamız, ortak ön yargıları ölçmenin ve onlarla mücadelenin önemini azaltmak için kullanılmamalıdır.

Her vakanın ortalama hüküm süresi, kolaylık olması açısından tablonun en alt sırasında gösteriliyor. Vakalar artan ciddiyet sıralamasına göre düzenlenmiş: A Vakası'nın ortalama hüküm süresi 1 yılken, P Vakası'nınki 15,3. On altı vakada ortalama hapis cezası ise 7 yıl.

Şimdi bütün yargıçların kusursuz adalet ölçme aleti kullandığı ve hükümlerin gürültüsüz olduğu kusursuz bir dünya hayal edin. Böyle bir dünyada Şekil 9 nasıl görünürdü? A Vakası sütunu aynı kalırdı, çünkü bütün yargıçlar A Vakası'ndaki sanığa tam bir yıllık aynı cezayı verirdi. Aynı şey diğer bütün sütunlar için de doğru olurdu. Her satırdaki sayılar ise, elbette, vakalar farklı olduğu için yine değişkenlik gösterirdi. Ancak her satır üstündeki

ve altındakiyle aynı olurdu. Vakalar arasındaki fark tablodaki tek değişkenlik kaynağı olurdu.

Ne yazık ki federal adalet dünyası kusursuz değil. Yargıçlar birbirinin aynısı değiller ve sütunlar içindeki değişkenlik, her vakadaki yargılarda gürültüyü işaret edecek kadar büyük. Hükümlerde olması gerekenden daha fazla çeşitlilik var ve çalışmanın amacı bu çeşitliliği analiz etmek.

Hüküm Piyangosu

Yukarıda tarif ettiğimiz ve bütün vakaların her yargıçtan aynı cezayı aldığı kusursuz dünya resminden başlayalım. Her sütuna birbirinin aynısı 208 sayı sıralanmış. Şimdi, her sütundan aşağı doğru inerek ve sağda solda sayıları değiştirerek, bazen ortalama hükme hapis süresi ekleyerek bazen bu hükümden hapis süresi çıkararak, gürültü katalım. Yapacağınız değişiklikler aynı olmayacağı için, sütun içinde değişkenlik yaratırlar. Bu değişkenlik gürültüdür.

Bu çalışmanın temel sonucu *her vakanın yargıları içinde* gözlenen büyük gürültü miktarıdır. Her vakadaki gürültü ölçümü, o vakaya verilen hapis cezasının standart sapmasıdır. Ortalama vaka için ortalama ceza 7 yıl, o ortalamanın etrafındaki standart sapma 3,4 yıldı.

Standart sapma terimine aşina olsanız da bu terimin somut bir tarifi faydalı olabilir. Rastgele iki yargıç seçtiğinizi ve bir vakadaki yargılarının farkını hesapladığınızı düşünün. Şimdi bunu bütün yargıç çiftleri ve vakalar için tekrarlayın ve sonuçların ortalamasını alın. Bu ölçüm, *ortalama mutlak fark*, size sanığın bir federal mahkeme salonunda karşı karşıya kaldığı piyangoya dair bir his vermiş olmalı. Yargıların normal bir şekilde dağıldığını varsayarak, standart sapmanın 1,128 katı ve bu, aynı vaka için rastgele seçilen iki hüküm arasındaki ortalama farkın 3,8 yıl olacağı anlamına geliyor. Bölüm 3'te bir sigorta şirketinin risk değerlendirme uzmanına ihtiyacı olan müşterinin karşılaştığı pi-

yangodan bahsetmiştik. Adli sanığın piyangosu, en hafif ifadeyle, sonuç açısından çok daha önemlidir.

Ortalama hüküm süresinin 7 yıl olduğu bir vakada, yargıçlar arasındaki 3,8 yıllık bir ortalama mutlak fark, rahatsız edici ve bizim açımızdan kabul edilemez bir sonuçtur. Bununla birlikte gerçek adalet idaresinde daha da fazla gürültü olduğundan şüphe duymamız için birçok neden vardır. İlk olarak, gürültü denetimindeki katılımcılar mukayese edilmeleri normalden çok daha kolay olan ve peş peşe sunulan yapay vakalarla uğraştılar. Gerçek hayat, tutarlılığın korunması için bu kadar çok destek sunmaz. İkinci olarak, bir mahkeme salonundaki yargıçlar burada sahip olduklarından çok daha fazla bilgi sahibidirler. Yeni bilgi, kati olmadığı sürece yargıçların birbirlerinden farklılık göstermeleri için daha fazla olanak yaratır. Bu nedenlerle sanıkların gerçek mahkeme salonlarında karşı karşıya kaldığı gürültünün burada gördüğümüzden daha büyük olduğundan şüpheleniyoruz.

Bazı Yargıçlar Katıdır: Yatay Gürültü

Analizin bir sonraki aşamasında, yazarlar gürültüyü ögelerine ayırdılar. Muhtemelen aklınıza ilk gelen gürültü yorumu –Yargıç Frankel'in de aklına geldiği gibi– gürültünün yargıçların katı hükümler vermeye yatkınlıkları arasındaki farklılığa bağlı olmasıdır. Herhangi bir savunma avukatının söyleyeceği gibi yargıçların kendilerine göre ünleri vardır: Ortalama bir yargıçtan daha katı olanlar, acımasız "cellat yargıçlar" olarak, ortalama bir yargıçtan daha merhametli olanlar ise "yufka yürekli yargıçlar" olarak anılır. Bu sapmalara *yatay hatalar* diyoruz. (Bir kez daha, burada hata ortalamadan sapma olarak tanımlanmıştır; ortalama yargıç yanılıyorsa, bir hata adaletsizliği bile giderebilir.)

Yatay hatalarda değişkenlik bütün yargı görevlerinde bulunabilir. Bazı şeflerin diğerlerinden daha cömert olduğu performans

değerlendirmeleri, bazı öngörücülerin diğerlerinden daha iyimser olduğu pazar payı öngörüleri ya da bazı ortopedistlerin diğerlerinden daha agresif olduğu bir durumda sırt ameliyatı tavsiyeleri, örnek olarak sayılabilir.

Şekil 9'daki her sıra tek bir yargıcın verdiği hükümleri gösteriyor. Her yargıcın verdiği ortalama hüküm (tablonun en sağındaki sütunda gösteriliyor) yargıcın katılık derecesinin bir ölçümüdür. Anlaşılacağı üzere yargıçlar bu boyutta çok büyük farklılıklar sergiliyor. En sağdaki sütundaki değerlerin standart sapması 2,4 yıl. Bu değişkenliğin adaletle hiçbir ilgisi yok. Aksine, sizin de şüphelenebileceğiniz gibi, ortalama hükümdeki farklar yargıçlar arasında başka özellik farklılıklarını –köken, eğitim, hayat tecrübeleri, politik görüşler, ön yargılar vs.– yansıtır. Araştırmacılar yargıçların genel olarak hüküm verme konusundaki tutumunu incelediler: Örneğin ceza vermenin başlıca amacının tecrit mi (suçluyu toplumdan ayırmak), rehabilitasyon mu, yoksa caydırıcılık mı olduğunu düşünüyorlardı? Araştırmacılar ana amacın rehabilitasyon olduğunu düşünen yargıçların, caydırıcılık ve tecritten yana olanlara göre daha kısa hapis cezaları ve daha uzun gözetim süresinde karar kıldığını gördüler. Ayrı bir şekilde Güney eyaletlerinde bulunan yargıçlar, ülkenin diğer yerlerindeki meslektaşlarına kıyasla hatırı sayılır derecede uzun cezalar veriyorlardı. Şaşırtıcı olmamakla birlikte muhafazakâr ideoloji de hükümlerin ciddiyetini etkiliyordu.

Genel sonuç, ortalama hüküm verme seviyesinin bir kişilik özelliği gibi çalıştığı şeklindeydi. Bu çalışmayı yargıçları iki ucunu çok sert ve çok merhametlinin tuttuğu bir ölçeğe yerleştirmek için kullanabilirsiniz; tıpkı bir kişilik testinin dışa dönüklük ve uyumluluk derecelerini ölçeceği gibi. Diğer özellikler misali hükümlerin katılığının genetik faktörlerle, hayat tecrübeleriyle ve kişiliğin diğer yönleriyle bağıntılı olmasını da bekleyebiliriz. Bunların hiçbirinin vaka ya da sanıkla bir ilgisi yoktur. *Yatay gürültü* terimini yar-

yargıçların ortalama yargılarının değişkenliği için kullanırız ve bu değişkenlik yatay hataların değişkenliğiyle aynıdır.

Yargıçlar Farklıdır: Örüntü Gürültüsü

Şekil 9'daki siyah okların gösterdiği gibi yatay gürültü 2,4, sistem gürültüsü ise 3,4 yıldır. Bu fark sistem gürültüsünün farklı yargıçlar arasındaki ortalama katılık farklarından ibaret olmadığını işaret ediyor. Gürültünün bu diğer ögesine *örüntü gürültüsü* diyeceğiz.

Örüntü gürültüsünü anlamak için Şekil 9'u tekrar düşünün ve rastgele seçilmiş bir hücreye –mesela C3'e– odaklanın. C Vakası'nın ortalama cezası sütunun en altında gösteriliyor; görebileceğiniz gibi 3,7 yıl. Şimdi 3 numaralı yargıcın bütün vakalarda verdiği ortalama cezayı bulmak için en sağdaki sütuna bakın. 5 yıl, yani büyük ortalamadan sadece 2,0 yıl az. Yargıçların katılığındaki farklılık 3. Sütun'daki gürültünün tek kaynağı olsaydı, C3 hücresindeki cezanın 3,7 – 2,0 = 1,7 yıl olacağını öngörürdünüz. Ancak C3'teki gerçek sayı 4 yıldır. Bu, 3 numaralı yargıcın o vakada hüküm verirken özellikle sert olduğuna işaret ediyor.

Aynı basit ve eklemeli mantık tablonun her sütunundaki her hükmü tahmin etmenize olanak sağlardı ama aslında çoğu hücrede basit modelden saplamalar bulurdunuz. Bir satır boyunca baktığınız zaman yargıçların bütün vakalardaki hükümlerinde eşit derecede katı olmadığını göreceksiniz bazılarında kendi kişisel ortalamalarından daha katı, bazılarında daha merhametliler. Bu artık sapmalara *örüntü hataları* diyoruz. Bu örüntü hatalarını tablonun her hücresine yazsaydınız, her yargıç için (satır) toplamlarının sıfır olduğunu ve her vaka için (sütun) toplamlarının yine sıfır olduğunu görürdünüz. Ancak örüntü hataları gürültüye katkıları açısından birbirlerini dengelemezler; çünkü gürültünün hesaplanması için bütün hücrelerdeki değerlerin karesi alınır.

Hüküm konusunda basit eklemeli modelin tutmadığını teyit etmenin daha kolay bir yolu var. Tabloda her sütunun altındaki ortalama hüküm süresinin soldan sağa doğru aralıksız şekilde yükseldiğini ama aynısının satırlar arasında doğru olmadığını görürsünüz. Örneğin 208 numaralı yargıç O Vakası'ndaki sanığa P Vakası'ndaki sanıktan daha yüksek ceza vermiş. Ayrı yargıçlar vakaları uygun buldukları hapis süresine göre sıralasalardı, sıralamaları aynı olmazdı.

Az önce saptadığımız çeşitlilik için *örüntü gürültüsü* terimini kullanıyoruz çünkü bu çeşitlilik yargıçların belli vakalara tavrında karmaşık bir örüntüyü yansıtır. Örneğin bir yargıç genel olarak ortalamadan daha sert ama beyaz yakalı suçlulara karşı nispeten daha merhametli olabilir. Bir diğeri hafif ceza verme eğilimindeyken, sanık tekrar tekrar suç işleyen sabıkalı olduğunda daha katı olabilir. Bir üçüncüsü ortalama katılığa yakın olabilir ama sanık sadece suç ortağı olduğunda daha anlayışlı, kurban yaşlıca bir insan olduğunda daha sert olabilir. (Örüntü gürültüsü tanımını okunabilirlik yararına kullanıyoruz. Örüntü gürültüsü için uygun istatistiki terim, "vakaya göre yargıç" şeklinde ifade edilen, *yargıç x vaka etkileşimi* olmalı. İstatistik eğitimi almış okuyuculardan yüklediğimiz çeviri sorumluluğu için özür dileriz.)

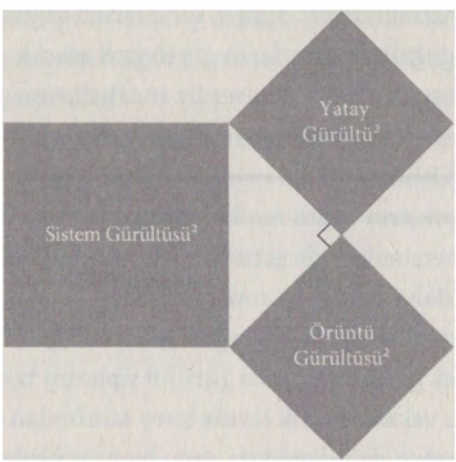
Adli yargı bağlamında vakalara verilen duruma özel tepkilerin bazıları yargıcın hüküm verme konusunda kişisel felsefesini yansıtıyor olabilir. Diğer tepkiler, bir sanığın özellikle nefret uyandıran bir suçluyu hatırlatması ya da belki de kızına benzemesi gibi yargıcın farkında bile olmadığı çağrışımların sonucu olabilir. Kökeni ne olursa olsun bu örüntüler, sadece tesadüfi değildir. Yargıcın aynı vakayı tekrar görmesi hâlinde tekrarlanmalarını bekleriz. Ancak örüntü gürültüsü, pratikte öngörülmesi güç olduğu için, zaten öngörülemez olan hüküm piyangosuna belirsizlik ekler. Çalışmanın yazarlarının ifade ettiği gibi "Yargıçlar arasında suç/suçlu özelliklerinin etkisi açısından örüntü farkları ek bir hüküm uyuşmazlığı biçimidir."

Sistem gürültüsünün yatay gürültü ve örüntü görüntüsü olarak çözümlenmesinin bir önceki bölümdeki, hatayı ön yargı ve gürültü olarak çözümleyen hata denklemiyle aynı mantığı izlediğini fark etmiş olabilirsiniz. Bu kez, denklem şu şekilde yazılabilir:

Sistem Gürültüsü² = Yatay Gürültü² + Örüntü Gürültüsü²

Bu ifade görsel olarak ilk Hata Denklemi'yle aynı tarzda temsil edilebilir (Şekil 10). Üçgenin iki kenarını eşit olarak temsil etmiştik. Bunun nedeni, hüküm vermeyle ilgili çalışmada, örüntü gürültüsü ve yatay gürültünün sistem gürültüsüne neredeyse eşit şekilde katkı sağlamasıydı.

Örüntü gürültüsü her yere nüfuz eder. Doktorların insanları hastaneye yatırıp yatırmamaya, şirketlerin kimi işe alacaklarına, avukatların hangi davaları mahkemeye taşıyacaklarına ya da Hollywood yapımcılarının hangi televizyon dizilerini yapacaklarına karar verişlerini düşünün. Bütün bu örneklerde farklı yargıçların, vakalara dair farklı puanlamalar üretmesi nedeniyle, örüntü gürültüsü olacaktır.



ŞEKİL 10: *Sistem gürültüsünün çözümlenmesi*

Gürültünün Bileşenleri

Örüntü gürültüsünü ele alışımız önemli bir karmaşıklığın üstünü örttü: rastgele hatanın olası katkısı.

Kronometre egzersizini hatırlayın. On saniyeyi tekrarlı olarak ölçmeye çalıştığınız zaman, sonuçlarınız bir turdan diğerine farklılık gösterdi; yani kişinin kendi içindeki değişkenliği sergilediniz. Aynı sebeple, yargıçlardan aynı vakaları başka bir durumda yeniden değerlendirmeleri istenseydi, söz konusu on altı davada aynı hükümleri vermezlerdi. Hatta göreceğimiz gibi ilk çalışma aynı haftanın farklı bir gününde yapılmış olsaydı da aynı hükümleri vermezlerdi. Eğer bir yargıcın keyfî kızına iyi bir şey olduğu ya da tuttuğu takım önceki gün kazandığı ya da güzel bir gün olduğu için yerindeyse, yargısı aksi takdirde olacağından daha merhametli olabilir. Bu kişinin içindeki değişkenlik az önce ele aldığımız istikrarlı kişiler arası farklılıklardan kavram olarak ayrıdır ama bu değişkenlik kaynaklarını birbirinden ayırt etmek zordur. Geçici etkilere bağlı olan bu değişkenliğe verdiğimiz isim *durum gürültüsü*dür.

Bu çalışmada durum gürültüsünü etkin bir şekilde yok saydık ve gürültü denetiminde yargıçların kendilerine özgü hüküm verme örüntülerini değişmez tavırların göstergesi olarak yorumlamayı seçtik. Bu varsayım elbette iyimserdir ancak durum gürültüsünün bu çalışmada büyük bir rol oynamadığına inanmak için bağımsız nedenler var. Çalışmaya katılan oldukça tecrübeli yargıçlar elbette suçların ve sanıkların çeşitli özelliklerinin önemi hakkında yerleşik fikirlerini de beraberlerinde getirdiler. Bir sonraki bölümde durum gürültüsünü daha detaylı olarak tartışacak ve sabit öge örüntü gürültüsünden nasıl ayrılabileceğini göstereceğiz.

Özetlemek gerekirse çeşitli gürültü tiplerini tartıştık. *Sistem gürültüsü* aynı vakalarda çok sayıda birey tarafından varılan yargılardaki istenmeyen değişkenliktir. Aynı bireyler birden fazla vakayı değerlendirdiğinde ayrılabilen başlıca iki ögesini saptadık.

- ❑ *Yatay gürültü* farklı yargıçların ortalama yargı seviyelerindeki değişkenliktir.
- ❑ *Örüntü gürültüsü* yargıçların belli vakalara verdikleri tepkilerdeki değişkenliktir.

Mevcut çalışmada yatay gürültü ve örüntü gürültüsü miktarları neredeyse eşitti. Ancak bizim örüntü gürültüsü olarak teşhis ettiğimiz bu bileşen, elbette rastgele hata olarak ele alınabilecek bir miktar *durum gürültüsü* de içermektedir.

Örnek olarak yargı sistemindeki bir gürültü denetimini kullandık ama aynı analiz iş hayatında, tıpta, hükûmette ya da başka yerlerde, her tür gürültü denetimine uygulanabilir. Hem yatay gürültü hem durum gürültüsü içeren örüntü gürültüsü, sistem gürültüsüne katkı sağlar ve ilerledikçe onlarla yine karşılaşacağız.

Gürültü Analizinden Bahsederken

"Yatay gürültü yargıçların farklı katılık seviyeleri sergilemesidir. Örüntü gürültüsü ise, hangi sanıkların daha katı ya da daha yumuşak muameleyi hak ettiği konusunda birbirleriyle anlaşmazlığa düşmeleridir. Ve örüntü gürültüsünün bir kısmını yargıçların kendileriyle görüş ayrılığına düştüğü durum gürültüsü oluşturur."

"Kusursuz bir dünyada, sanıklar adaletle karşı karşıya gelirdi; bizimkinde ise, gürültülü bir sistemle karşı karşıya kalıyorlar."

BÖLÜM 7

Durum Gürültüsü

Profesyonel bir basketbolcu serbest atışa hazırlanıyor. Faul çizgisinde duruyor. Konsantre oluyor ve atışını yapıyor. Bu sayısız kez çalıştığı net bir hareket dizisi. Atışı yapabilecek mi?

Bilmiyoruz, o da bilmiyor. Ulusal Basketbol Birliği'nde (NBA) oyuncular, tipik olarak atışların dörtte üçünü sayıya çeviriyorlar. Tahmin edileceği gibi bazı oyuncular diğerlerinden daha iyiler ama hiçbir oyuncu atışlarının %100'ünü sayıya dönüştüremiyor. Bütün zamanların en iyi oyuncuları serbest atışlarının %90'ından biraz fazlasında sayıya ulaşmıştır. (Bu kitabın yazıldığı zamanda bu sporcular Stephen "Steph" Curry, Steve Nash ve Mark Price idi.) Bütün zamanların en kötüleri ise %50 civarında. (Örneğin büyük Shaquille O'Neal atışlarının sadece %53'ünü isabet ettirmiştir.) Pota her zaman tam olarak 3 metre yüksekte ve 4,6 metre uzakta, top her zaman 623 gram olsa da sayı kaydetmek için gereken kesin jest sıralaması her zaman çok kolay değildir. Sadece oyuncular arasında değil, oyuncunun kendi içinde de değişkenlik beklenir. Serbest atış bir tür piyangodur. Oyuncunun O'Neal yerine Curry olması durumunda başarı şansı daha yüksektir ama yine de piyangodur.

Bu değişkenlik nereden gelir? Faul çizgisindeki oyuncuyu sayısız faktörün etkileyebileceğini biliyoruz: Uzun bir maçın yorgunluğu, başa baş bir skorun zihinsel baskısı, ev sahibi takımın taraftarlarının tezahüratları ya da rakip takımın taraftarlarının moral bozucu sesleri. Curry ya da Nash gibi biri atışı ıskaladığında aklımıza bu açıklamalardan biri gelir. Ancak aslında, bu faktörlerin oynadığı kesin rolü bilmemiz çok mümkün değildir. Atış yapan bir oyuncunun performansındaki değişkenlik gürültünün bir türüdür.

İkinci Piyango

Serbest atışlardaki ya da diğer fiziksel süreçlerdeki değişkenlik sürpriz değildir. Vücutlarımızdaki değişkenliğe alışığızdır; nabzımız, kan basıncımız, reflekslerimiz, ses tonumuz ve ellerimizdeki titreme farklı zamanlarda farklılık gösterir. Aynısını yaratmak için ne kadar çabalasak da imzamız her çekte çok az farklılaşır.

Zihinlerimizin değişkenliğini gözlemlemek daha az kolaydır. Elbette yeni bilgi bile olmadan fikir değiştirme tecrübesini hepimiz yaşamışızdır. Dün gece bizi kahkahalarla güldüren film, şimdi vasat ve unutulabilir geliyordur. Dün katı bir şekilde yargıladığımız bir insan bugün hoşgörümüzü hak ediyor gibi gelebilir. Hoşumuza gitmeyen ya da anlamadığımız bir argüman, sonradan kafamıza yatıp gerekli görünebilir. Ancak bu örneklerin düşündürdüğü gibi genelde bu tür değişiklikleri nispeten önemsiz ve büyük ölçüde öznel meselelerle bağdaştırırız.

Gerçekte fikirlerimiz ortada görünür bir neden olmadan değişir. Bu nokta profesyonel uzmanların dikkatli ve çok düşünülmüş yargıları için bile geçerlidir. Örneğin aynı vaka önlerine iki kez konduğunda, aynı doktorlardan önemli ölçüde farklı teşhisler almak yaygın bir durumdur (bakınız Bölüm 22). Büyük bir Amerikan şarap yarışmasında, şarap uzmanları aynı şarapları iki kez tattıkları zaman, şarapların sadece %18'ine (çoğunlukla en kötü

olanlara) aynı notları verdiler. Bir adli tıp uzmanı aynı parmak izlerini sadece birkaç hafta arayla incelediği zaman farklı sonuçlara varabilir (bakınız Bölüm 20). Uzman yazılım danışmanları iki ayrı zamanda, aynı görevin tamamlanma süresi için çok farklı tahminlerde bulunabilir. Basit bir ifadeyle aynı topu iki kez aynı şekilde atamayan bir basketbol oyuncusu gibi, aynı olgularla farklı durumlarda karşı karşıya kaldığımızda, her zaman birbirine eş yargılar üretmeyiz.

Bir risk değerlendirme uzmanını, bir yargıcı ya da bir doktoru seçen süreci sistem gürültüsü yaratan bir piyango olarak tarif etmiştik. Durum gürültüsü ikinci bir piyangonun ürünüdür. Bu piyango profesyonel kişinin bir yargıda bulunduğu anı, ruh hâlini, zihinde tazeliğini koruyan vakalar dizisini ve durumun diğer sayısız özelliklerini seçer. Bu ikinci piyango genellikle ilkinden çok daha soyut kalır. Örneğin birinci piyangonun nasıl farklı bir risk değerlendirme uzmanı seçebileceğini görebiliriz ama seçilen risk değerlendirme uzmanının gerçek tepkilerinin alternatifleri soyut karşıolgusallardır. Sadece gerçekleşen yargının bir olasılıklar bulutunun içinden seçildiğini biliriz. Durum gürültüsü bu görünmeyen değişkenler arasındaki değişkenliktir.

Durum Gürültüsünü Ölçmek

Durum gürültüsünü ölçmek, varlığıyla aynı nedenden ötürü, kolay değildir; varlığı anlaşıldığı zaman bizi genellikle şaşırtır. İnsanlar özenle düşünülmüş profesyonel bir görüş oluşturdukları zaman, onu bakış açılarını haklı gösteren nedenlerle bağdaştırırlar. Yargılarını açıklamaları için üstelendiğinde, onları genellikle ikna edici buldukları argümanlarla savunurlar. Aynı sorun önlerine ikinci kez konulur ve sorunu tanırlarsa hem çabayı en aza indirmek hem tutarlılığı sağlamak için bir önceki cevabı tekrar üretirler. Öğretmenlik mesleğinden şu örneği düşünün: Bir öğretmen, bir

öğrencinin denemesine kusursuz bir not verir ve sonra aynı denemeyi bir hafta sonra verilen ilk notu görerek yeniden okursa, farklı bir not verme ihtimali çok düşüktür.

Bu nedenle vakaların kolayca hatırda kalır olduğu örneklerde, durum gürültüsünün direkt ölçümü zordur. Örneğin bir risk değerlendirme uzmanına ya da bir yargıca daha önce karar vardıkları bir vakayı gösterirseniz, muhtemelen onu tanır ve bir önceki yargılarını tekrar ederler. Profesyonel yargıda değişkenlik üzerine yapılan bir araştırma (teknik olarak *test-yeniden test güvenilirliği* ya da kısaca *güvenilirlik* olarak bilinir) uzmanların aynı oturumda aynı yargıya iki kez vardığı pek çok çalışmayı içeriyor. Uzmanların doğru cevap olma eğiliminde olmasına şaşmamalı.

Yukarıda bahsettiğimiz deneyler bu hususu uzmanların tanımayacağı uyarıcılar kullanarak bertaraf ettiler. Şarap jürileri bir kör tadıma katıldılar. Parmak izi uzmanlarına daha önce gördükleri iz çiftleri gösterildi ve yazılım uzmanlarına daha önce üstünde çalıştıkları görevlerle ilgili sorular soruldu. Ancak aradan birkaç hafta ya da ay geçmesi beklendi ve uzmanlara bu örnekleri daha önce incelediklerini söylenmedi.

Durum gürültüsünün varlığını teyit etmenin daha dolambaçlı bir yolu daha var: büyük veriler ve ekonometri yöntemleri kullanmak. Geçmiş profesyonel kararlardan çok büyük bir örnek havuzu kullanılabilir olduğunda, analizciler bazen bu kararların nasıl alındığını merak edip duruma özgülük, saat ya da hava sıcaklığı gibi alakasız faktörlerden etkilenip etkilenmediğini kontrol edebilirler. Bu tür ilişkisiz faktörlerin yargılar üstündeki istatistiki açıdan önemli etkileri durum gürültüsünün kanıtıdır. Gerçekçi bir bakış açısıyla söylemek gerekirse, konuyla alakasız bütün durum gürültüsü kaynaklarını keşfetmeyi umut edemeyiz ancak bulunabilenler, bu kaynakların büyük çeşitliliğini gösterirler. Durum gürültüsünü kontrol etmek için, onu üreten mekanizmaları anlamaya çalışmalıyız.

Bir Kişi de Kalabalıktır

Şu soruyu bir düşünün: Dünyadaki havaalanlarının yüzde kaçı ABD'dedir? Düşünürken, muhtemelen aklınıza bir cevap geldi. Ama yaşınızı ya da telefon numaranızı hatırlayacağınız şekilde değil. Ürettiğiniz sayının bir tahmin olduğunun farkındasınız. Bu rastgele bir sayı değil; %1 ya da %99 yanlış cevaplar olurdu. Ancak bulduğunuz sayı üstünü çizmeyeceğiniz olasılıklar aralığı içinden sadece biri. Biri cevabınıza %1 puan eklese ya da cevabınızdan %1 puan çıkarsa, muhtemelen ortaya çıkan tahminin sizinkinden daha az makul olduğunu düşünmezsiniz. (Merak ediyorsanız, doğru cevap %32.)

Edward Vul ve Harold Pashler adında iki araştırmacı, insanlardan bu (ve benzer daha pek çok) soruyu bir kez değil, iki kez cevaplamalarını istemeyi akıl ettiler. Deneklere ilk seferinde bir tahminde daha bulunacakları söylenmedi. Vul ve Pashler'in hipotezi iki cevabın ortalamasının her iki cevabın tek başına olduğundan daha doğru olacağı şeklindeydi.

Veri onları haklı çıkardı. İlk tahmin, genel olarak gerçeğe ikincisinden daha yakındı ama en iyi tahmin iki tahminin ortalamasından geldi.

Vul ve Pashler, *kalabalıkların bilgeliği etkisi* olarak bilinen bir fenomenden (farklı insanların bağımsız yargılarının ortalamasını almak genellikle gerçek değere yakınlığı artırır) ilham almışlardı. 1907'de, Darwin'in kuzeni ve birçok disiplinle uğraşan bir bilim insanı olan Francis Galton bir taşra panayırında 787 köylüden ödül olarak verilecek bir öküzün ağırlığını tahmin etmelerini istedi. Köylülerin hiçbiri öküzün 543 kilo olan gerçek ağırlığını tahmin edemedi ancak tahminlerinin ortalaması 544 kiloydu, yani sadece bir kilo farklıydı ve medyan (547) da çok yakındı. Köylüler bireysel tahminleri hayli gürültülü olmasına rağmen yanlı olmamaları açısından "bilge bir kalabalık" oluşturmuştu. Bu saptama Galton'ı şaşırtmıştı: Sıradan insanların yargısına çok saygı duymazdı ve

buna rağmen, çalışmasının sonuçlarının "demokratik bir yargının güvenilirliğine beklenenden daha fazla itibar kazandırdığı"nı öne sürmüştü.

Benzer sonuçlara yüzlerce durumda rastlanmıştır. Elbette sorular cevaba sadece uzmanların yaklaşabileceği kadar zor olduğunda, kalabalıklar doğru sonuca illa ulaşamayacaktır. Ancak insanlardan örneğin şeffaf bir kavanozun içindeki şekerlerin sayısını, şehirlerinin bir hafta sonraki hava sıcaklığını ya da bir eyaletteki iki şehir arasındaki mesafeyi tahmin etmeleri istendiğinde, çok sayıdaki insanın cevaplarının ortalamasının doğruya yakın olması çok olasıdır. Nedeni temel istatistiktir: Çok sayıda bağımsız yargının (ya da ölçümün) ortalamasını almak, yeni ve bireysel yargılardan daha az ön yargılı olmasa da daha az gürültülü bir yargı üretir.

Vul ve Pashler aynı etkinin durum gürültüsüne de uzanıp uzanmayacağını merak ettiler: Farklı insanların tahminlerini birleştirdiğimizde olduğu gibi, aynı insanın iki tahmini birleştirilerek gerçeğe daha fazla yaklaşılabilir miydi? Ortaya çıkardıkları gibi, cevap evet. Vul ve Pashler bu bulguya çağrışımsal bir ad verdiler: *içimizdeki kalabalık.*

Aynı kişinin iki tahmininin ortalamasını almak, yargıyı bağımsız bir ikinci görüş aramak kadar iyileştirmiyor. Vul ve Pashler'in ifadesiyle "Aynı soruyu kendinize iki kez sorarak, bir başkasından alacağınız ikinci bir görüşten edineceğinizin ancak onda birini elde edebilirsiniz." Bu büyük bir ilerleme değil. Ancak ikinci bir tahminde bulunmayı bekleyerek etkiyi çok daha büyütebilirsiniz. Vul ve Pashler süjelerine aynı soruyu sormak için aradan üç hafta geçmesini bekledikleri zaman, sağlanan fayda ikinci bir kişinin görüşünün değerinin üçte birine yükseldi. Ek bir bilgi ya da dışarıdan yardım gerektirmeyen bir teknik için hiç fena değil. Ayrıca bu sonucun karar alıcılara çok eski bir tavsiyeyi teyit ettiği de kesin: "Üstüne bir gece uyuyun ve ertesi sabah tekrar düşünün."

Vul ve Pashler'den bağımsız ancak yaklaşık aynı zamanda çalışan iki Alman araştırmacısı Stefan Herzog ve Ralph Hertwig, aynı prensibin farklı bir uygulamasına ulaştılar. Deneklerden ikinci bir tahmin üretmelerini istemek yerine, onları ilkinden olabildiğince farklı –hâlâ makul olmak üzere– bir tahmin üretmeye teşvik ettiler. Bu talep deneklerin ilk seferinde üstünde durmadıkları bilgileri aktif olarak düşünmelerini gerektirdi. Katılımcılara verilen talimatlar şöyleydi:

> *Önce ilk tahmininizin hedefi ıskaladığını varsayın. İkinci olarak bunun neden olmuş olabileceği konusunda birkaç sebep düşünün. Hangi varsayımlar ve değerlendirmeler yanlış olabilir? Üçüncü olarak bu yeni değerlendirmeler neyi ima ediyor? İlk tahmin çok mu yüksekti, yoksa çok mu düşüktü? Dördüncü olarak bu yeni perspektife dayanarak ikinci, alternatif bir tahmin yapın.*

Vul ve Pashler gibi Herzog ve Hertwig de bu şekilde üretilen iki tahminin ortalamasını aldılar. *Diyalektik yeniden örnekleme* adını verdikleri teknikleri, doğruya yaklaşmak konusunda basit ilkini takiben ikinci bir tahminde bulunma talebinden daha büyük iyileşmeler üretti. Katılımcılar kendilerini soruyu yeni bir ışıkta değerlendirmeye zorladıkları için kendilerinin daha farklı bir versiyonunu örnek aldılar: "içimizdeki kalabalığın" birbirinden çok uzak olan iki "üyesi". Sonuç olarak ortalamaları gerçeğin daha doğru bir tahminini üretti. Peş peşe yapılan iki diyalektik tahminin doğruluğa getirisi ikinci bir kişinin görüşünün neredeyse yarı değerindeydi.

Herzog ve Hertwig'in özetledikleri hâliyle karar alıcılar için sonuç, prosedürler arasında basit bir tercihtir: Başkalarından bağımsız görüş alabiliyorsanız, alın: Kalabalıkların gerçek bilgeliğinin yargınızı iyileştirmesi kuvvetle muhtemel. Ama alamıyorsanız bir

"iç kalabalık" yaratmak için aynı yargıyı kendiniz ikinci bir kez değerlendirin. Bunu aradan biraz zaman geçtikten sonra –ve ilk görüşünüzle aranıza mesafe koyarak– da yapabilirsiniz, soruna başka bir bakış açısı bulmak için kendinizle aktif olarak tartışmaya çalışarak da. Sonuç olarak kalabalık türünden bağımsız olarak, tahminlerinizden birine daha fazla ağırlık vermek için çok güçlü nedenleriniz olmadığı sürece, en iyi çare iki tahminin ortalamalarını almak olacaktır.

Pratik tavsiyenin ötesinde bu araştırma hattı, yargı konusunda önemli bir içgörüyü doğruluyor. Vul ve Pashler'in ifadesiyle "Bir denek tarafından verilen cevaplar, deneğin sahip olduğu bütün bilgiye dayanarak belirleyici bir şekilde seçilmek yerine, içsel bir olasılık dağılımından örnek alınır." Bu gözlem, ABD'deki havaalanlarıyla ilgili soruyu yanıtlarken yaşadığınız tecrübeyi yansıtıyor. İlk cevabınız bütün bilginizi, hatta en iyi kısmını bile yakalayamadı. Cevabınız zihninizin ürettiği olası cevaplar bulutundan sadece bir noktaydı. Aynı sorun için aynı kişi tarafından varılan yargılarda gözlemlediğimiz değişkenlik uzmanlık gerektiren birkaç sorunda gözlenen bir tesadüf değil: Durum gürültüsü bütün yargılarımızı, her zaman etkiler.

Durum Gürültüsünün Kaynakları

Hepimizin fark ettiği en az bir durum gürültüsü kaynağı vardır: ruh hâli. Hepimiz kendi yargılarımızın nasıl hissettiğimize göre değiştiğini tecrübe etmişizdir ve başkalarının yargılarının da ruh hâllerine göre değişkenlik sergilediğinin farkındayızdır.

Ruh hâlinin yargı üstündeki etkisi çok sayıda psikolojik araştırmanın konusu olmuştur. İnsanları geçici olarak mutlu etmenin ya da üzmenin ve ruh hâllerinin kışkırtılmasının ardından, yargı ve kararlarının değişkenliğini ölçmek çok kolaydır. Araştırmacılar bunu yapmak için çok çeşitli teknikler kullanırlar. Örneğin zaman

zaman katılımcılardan mutlu ya da üzücü bir anıyı anlatan bir paragraf yazmaları istenir. Katılımcılar bazen sadece komik ya da acıklı bir filmden alınmış bir video izlerler.

Sayısız psikolog onlarca yılını ruh hâli manipülasyonunun etkilerini soruşturmaya harcamıştır. Belki de aralarında en üretken olan Avustralyalı psikolog Joseph Forgas'tır. Forgas ruh hâli konusunda yaklaşık yüz bilimsel makale yayımlamıştır.

Forgas'ın araştırmalarının bazıları zaten düşündüğünüz bir şeyi teyit eder: İyi bir ruh hâlinde olan insanlar genellikle daha pozitiftir. Hüzünlülerden çok mutlu anıları hatırlamak onlara daha kolay gelir, insanlara karşı daha onaylayıcıdırlar, daha cömert ve yardımseverdirler, gibi. Olumsuz ruh hâli ise tam tersi etkiler yaratır. Forgas'ın yazdığı gibi "İyi bir ruh hâlindeki bir insan tarafından sevecen olarak görülen bir gülümseme, gözlemci olumsuz bir ruh hâlinde olduğunda, sakil bulunabilir. Kişi iyi bir ruh hâlindeyken hava durumunu tartışmak temkinli bir tavır gibi görünebilirken, kötü ruh hâlinde olduğunda sıkıcı gelebilir."

"Başka bir deyişle ruh hâlinin düşünceleriniz üstünde ölçülebilir bir etkisi vardır. Ortamınızda neyi fark edeceğiniz, hafızanızdan neyi geri çağıracağınız, bu sinyallere nasıl anlam yüklediğiniz gibi. Ancak ruh hâlinin şaşırtıcı bir etkisi daha var: Aynı zamanda *nasıl* düşündüğünüzü de etkiler. Ve burada, etkiler aklınıza gelebilecek türden değildir. İyi bir ruh hâlinde olmak, karışık bir lütuftur ve kötü ruh hâllerinin de iyi bir tarafı vardır. Farklı ruh hâllerinin maliyetleri ve getirileri duruma özeldir.

Örneğin pazarlık durumunda iyi ruh hâli yardımcı olur. İyi bir ruh hâlindeki insanlar daha iş birlikçidir ve alışverişin yolunu açar. Mutsuz pazarlıkçılardan daha iyi sonuçlar almaya daha yakındırlar. Elbette başarılı pazarlıkçılar insanları da mutlu ederler ama bu deneylerde, ruh hâline pazarlıkta yaşananlar neden olmuyor; iyi ruh hâli pazarlığın öncesinde kışkırtılıyor. Ayrıca pazarlık sırasında iyi bir ruh hâlinden öfkeli bir ruh hâline geçenler de

genellikle iyi sonuçlar alıyorlar: İnatçı bir muhatapla karşı karşıya olduğunuzda hatırlanacak bir şey!

Öte yandan iyi bir ruh hâli ilk izlenimlerimizi kafa tutmadan doğru olarak kabul etme olasılığımızı artırır. Forgas'ın çalışmalarından birinde, katılımcılar yazarın bir fotoğrafının eklendiği kısa bir felsefi deneme okudular. Okuyuculardan bazıları klişe bir felsefe profesörü gördü: erkek, orta yaşlı, gözlüklü. Diğerleri genç bir kadın gördü. Tahmin edebileceğiniz gibi bu okuyucuların klişelere karşı savunmasızlığıyla ilgili bir testti: İnsanlar orta yaşlı bir erkeğe atfedildiğinde denemeyi, genç bir kadının yazdığına inandıkları zamankinden daha olumlu mu değerlendireceklerdi? Elbette, öyle oldu. Ama asıl önemli olan, farkın iyi ruh hâli şartında daha büyük olmasıydı. İyi bir ruh hâlindeki insanlar, ön yargılarının düşüncelerini etkilemesine izin vermeye daha yatkın olurlar.

Başka çalışmalar da ruh hâlinin kolay kandırılma üstündeki etkisini test etti. Gordon Pennycook ve meslektaşları insanların popüler guruların deyişlerinden rastgele seçilen isim ve fiillerin dil bilgisi açısından doğru cümlelerde bir araya getirilmesiyle üretilmiş anlamsız, sözde derin beyanlara tepkileri üstüne sayısız çalışma yaptılar. Örneğin: "Bütünlük sonsuz fenomenleri susturur." ya da "Gizli anlam benzersiz soyut güzelliği dönüştürür." gibi. Bu tür beyanlarla hemfikir olma eğilimi boktanlığa *açıklık* olarak bilinen bir özelliktir. (*Boktanlık* kelimesi Princeton Üniversitesi'nden Harry Frankfurt adlı bir filozofun boktanlığı diğer yanlış betimleme türlerinden ayırdığı *Boktanlık Üstüne* adlı aydınlatıcı kitabının yayımlanmasından bu yana bir tür teknik terime dönüştü.)

Elbette bazı insanlar boktanlığa diğerlerinden daha açıktır. "Doğru ve anlamlı olarak sunulan ama aslında kof olan, görüntüde etkileyici savlardan" etkilenebilirler. Ancak burada da bu saflık sadece kalıcı, değişmez yaradılışların bir işlevi değildir. İyi ruh hâlini kışkırtmak insanları boktanlığı almaya ve genel olarak kolay kandırılmaya daha açık hâle sokar: Bu ruh hâlindeki insan-

lar, hileyi fark etmeye ya da yanıltıcı bilgiyi saptamaya daha az yatkın olurlar. Tam tersi şekilde, yanıltıcı bilgiye maruz bırakılan görgü tanıkları da kötü bir ruh hâlinde olduklarında, yanıltıcı bilgiyi umursamamaya –ve yanlış tanıklıktan kaçınmaya– daha yatkın olurlar.

Ahlaki yargılar bile ruh hâlinden güçlü bir şekilde etkilenir. Bir çalışmada araştırmacılar denekleri ahlak felsefesinin klasik bir problemi olan üst geçit problemiyle karşı karşıya bıraktılar. Bu düşünce deneyinde, beş kişi kontrolden çıkmış bir tramvay tarafından öldürülmek üzereler. Deneklerden kendilerini tramvayın birazdan altından geçeceği üst geçitte dururken hayal etmeleri isteniyor. Vücuduyla tramvayı durdursun diye iri yapılı bir adamı üst geçitten aşağı rayların üstüne itip itmemeye karar vermeliler. Onlara bunu yapmaları hâlinde iri yapılı adamın öleceği ama aşağıdaki beş kişinin kurtulacağı söyleniyor.

Üst geçit problemi ahlaki akıl yürütmeye yaklaşımlar arasındaki çelişkiyi temsil ediyor. İngiliz filozof Jeremy Bentham'la anılan faydacı hesaplama, bir kişinin hayatını kaybetmesinin beş kişininkine yeğ olduğunu öne sürer. Immanuel Kant'la bağdaştırılan deontolojik etik, başkalarını kurtarmak için bile olsa can almayı yasaklar. Üst geçit problemi kişisel duygunun çarpıcı bir ögesini açıkça içerir: Fiziksel olarak bir adamı bir köprüden yaklaşan bir tramvayın yoluna itmek özellikle tiksinti uyandıran bir eylemdir. Faydacı tercihi yaparak adamı üst geçitten itmek insanların bir yabancıya karşı fiziksel şiddet içeren bir eylemden kaçınma dürtülerini aşmalarını gerektirir. Genellikle insanların çok az kısmı (bu çalışmada onda birden de azı) bunu yapacağını söyler.

Bununla birlikte denekler, beş dakikalık bir video izletilerek olumlu bir ruh hâline sokulduklarında adamı köprüden iteceklerini söylemeye yatkınlıkları üç kat arttı. "Öldürmeyeceksin" emrini mutlak bir prensip olarak görüp görmediğimiz ya da beş yabancıyı kurtarmak için bir yabancıyı öldürmeye istekli olup olmamamız

en derin değerlerimizi yansıtmalıdır. Buna rağmen tercihimiz az önce hangi videoyu izlediğimize bağlı görünüyor.

Bu ruh hâli çalışmalarını detaylı olarak anlattık çünkü önemli bir gerçeği vurgulamamız gerekiyor: *Her zaman aynı insan değilsiniz.* Ruh hâliniz değiştikçe (elbette bu farkında olduğunuz bir şey) bilişsel mekanizmanızın bazı özellikleri de onunla birlikte değişkenlik gösterir (bunun ise tamamen farkında *değilsiniz*). Önünüze karmaşık bir yargı problemi konduğunda o andaki ruh hâliniz, ruh hâlinizin böyle bir etkisi olmadığına inandığınızda ve hatta bulduğunuz cevabı kendinizden emin bir şekilde gerekçelendirebildiğinizde bile, soruna yaklaşımınızı ve varacağınız sonuçları değiştirebilir. Kısaca gürültülüsünüz.

Daha pek çok tesadüfi olay, yargılarda durum gürültüsüne neden olur. Profesyonel yargıları etkilememesi gereken ama etkileyen dış kaynaklı faktörler arasında başlıca iki şüpheli, stres ve yorgunluktur. Örneğin yaklaşık yedi yüz bin birinci basamak sağlık bakım servisinde yapılan bir çalışma, doktorların uzun bir günün sonunda opioid ilaçlar yazmaya daha yatkın olduğunu gösterdi. Elbette saat 16.00'da randevusu olan bir hastanın sabah 09.00'da gelen bir hastadan daha büyük bir acı çekiyor olması için bir neden yok. Doktorun programında sarkma yaşıyor olması da reçete kararlarını etkilememeli. Ayrıca, streoid içermeyen antienflamatuar ilaçlar gibi diğer ağrı kesicilerin reçete edilmesi ya da fizik tedavi tavsiyeleri benzer örüntüler sergilemiyor. Doktorlar zaman baskısı alında olduklarında, ciddi olumsuzluklarına rağmen, sorunu hızlı bir şekilde çözecek bir tercih yapmaya daha yatkın oluyorlar. Diğer çalışmalar, günün sonuna doğru, doktorların antibiyotik yazmaya daha fazla, grip aşısı yazmaya daha az yatkın olduğunu gösteriyor.

Hava durumunun bile profesyonel yargılar üstünde ölçülebilir bir etkisi vardır. Bu yargılara genellikle klimalı odalarda varıldığı için, hava durumunun etkisi muhtemelen ruh hâli üzerinden gerçekleşiyor (yani hava durumu kararları doğrudan etkilemiyor ama

karar alıcının ruh hâlini, ruh hâli de nasıl bir karar verileceğini etkiliyor). Kötü hava durumu daha iyi hafızayla ilişkilendirilir, adli yargı hükümleri hava sıcakken daha katı olur, borsanın performansı güneş ışığından etkilenir. Bazı durumlarda, hava durumunun etkisi daha az barizdir. Uri Simonsohn üniversite kabul bürosu memurlarının öğrencilerin akademik özelliklerine bulutlu günlerde daha fazla dikkat ettiklerini ve güneşli günlerde akademik olmayan özelliklerine karşı daha duyarlı olduklarını gösterdi. Bu bulgularını paylaştığı makalenin başlığı hatırda yer edecek türden: "Bulutlar İnekleri İyi Gösteriyor."

Yargıdaki rastgele değişkenliğin bir diğer kaynağı da vakaların incelendiği sıradır. Bir kişi bir vakayı değerlendirirken, o vakanın hemen öncesinde verilen kararlar üstü örtülü bir referans iskeleti görevi görür. Yargıçlar, kredi memurları, beyzbol hakemleri gibi peş peşe kararlar alan profesyoneller bir tür dengeyi korumaya meyilli görünüyorlar: Aynı yönde giden bir dizi karardan sonra, katı bir şekilde gerekçelendirilebilecekleri aksi yönde kararlar almaya daha yatkın olurlar. Sonuç olarak, hatalar (ve haksızlıklar) kaçınılmazdır. Örneğin ABD'de önceki iki dosya onaylandığında, sığınma yargıçlarının sığınma hakkı tanıma ihtimali %19 azalır. Kişinin kredi talebi ondan önceki iki kişinin geri çevrilmesi hâlinde artar ama önceki iki kişiye kredi verilmesi hâlinde, aynı kişi ret cevabı alabilir. Bu davranış *kumarbaz yanılgısı* olarak bilinen bilişsel bir yanlılığı yansıtır: Aynı yöndeki durumların tesadüfen üst üste gelebileceği olasılığını hafife almaya meyilliyiz.

Durum Gürültüsünü Ölçmek

Durum gürültüsü toplam sistem gürültüsüne kıyasla ne kadar büyüktür? Her ne kadar bütün durumlara uyan tek bir sayı olmasa da genel bir kural kendini gösteriyor: Boyutları açısından,

bu bölümde tarif ettiğimiz etkiler bireylerin yargı seviyeleri ve örüntüleri arasındaki sabit farklardan daha küçüktür.

Daha önce değindiğimiz gibi sığınma talebinde bulunan bir kişinin ABD'ye kabul edilme olasılığı, duruşmasının aynı yargıcın olumlu karar verdiği iki başarılı davanın hemen arkasından gelmesi durumunda %19 azalıyor. Bu değişkenlik elbette can sıkıcıdır. Ancak yargıçlar arasındaki değişkenlikle kıyaslandığında sönük kalır: Jaya Ramji-Nogales ve meslektaşları bir Miami mahkemesinde bir yargıç başvuru sahiplerinin %88'ine sığınma hakkı tanırken, bir diğerinin sadece %5'inin başvurularını kabul ettiğini ortaya çıkardılar. (Bu bir gürültü denetimi değil, gerçek veridir. Yani başvuru sahipleri farklıydı ama yarı rastgele dağıtılmışlardı ve yazarlar başvuru sahiplerinin geldikleri ülkenin uyuşmazlıklarını açıklamadığından emin oldular.) Bu tür farklılıklar söz konusuyken, bu sayılardan birini %19 azaltmak çok önemli görünmüyordu.

Benzer bir şekilde parmak izi uzmanları ve doktorlar bazen kendi kendileriyle görüş ayrılığı yaşarlar ama bu başkalarıyla yaşadıkları anlaşmazlıklara göre çok daha nadirdir. İncelediğimiz ve toplam sistem gürültüsünde durum gürültüsünün payının ölçülemediği her vakada, durum gürültüsü sonuca bireyler arasındaki farklardan daha küçük bir katkı sağlıyordu.

Ya da daha farklı bir ifadeyle her zaman aynı insan değilsiniz ama zaman ilerlerken sandığınızdan daha az tutarlı kalırsınız. Ancak bir şekilde iç rahatlatıcı olan, dünkü hâlinize bugün başka bir insana benzediğinizden daha fazla benzemenizdir.

Durum Gürültüsü, İç Nedenler

Ruh hâli, yorgunluk, sekans etkileri: Aynı örnek için aynı kişinin yargısında istenmeyen farklılıkların oluşmasını pek çok faktör tetikleyebilir. Kararlar üstünde etkili olan bütün alakasız dış faktörlerin bilindiği ve kontrol edildiği bir ortam yaratmayı umabiliriz. En

azından teoride böyle bir ortam durum gürültüsünü azaltır. Öte yandan muhtemelen bu ortam bile durum gürültüsünü tamamen ortadan kaldırmaya yetmez.

Michael Kahana ve Pensilvanya Üniversitesi'nden meslektaşları hafıza performansı üstüne çalışmalar yapıyorlar. (Bizim tanımımıza göre hafıza bir yargı görevi değildir ancak şartların katı bir şekilde kontrol edilebildiği ve performanstaki farklılıkların kolayca ölçülebildiği bir bilişsel görevdir.) Bir çalışmada yetmiş dokuz denekten hafıza performanslarının olağanüstü ayrıntılı analizine katılmalarını istediler. Denekler ayrı günlerde, her seferinde yirmi dört kelimeden oluşan yirmi dört farklı listeden sözcükleri akıllarında tutmalarının istendiği yirmi üç oturuma katıldılar. Akılda kalan sözcüklerin yüzdesi hafıza performansını tanımlıyordu.

Kahana ve arkadaşları denekler arasındaki farklardan çok, her deneğin performansındaki değişkenlik kestiricileri arasındaki farklarla ilgiliydiler. Performans deneğin kendini ne kadar uyanık hissettiğinden etkilenecek miydi? Bir önceki gece aldığı uyku miktarından? Günün saatinden? Performans pratik sayesinde bir oturumdan diğerine iyileşebilir miydi? Denek yoruldukça ya da sıkıldıkça, her oturum içinde bozulabilir miydi? Bazı sözcük listelerinin ezberlenmesinin diğerlerinden daha kolay olduğu ortaya çıkabilir miydi?

Bütün bu soruların cevabı evetti ama çok büyük bir farkla değil. Bütün bu kestiricileri kapsayan bir model belli bir deneğin performansındaki değişikliğin sadece %11'ini açıklıyordu. Araştırmacıların ifadesiyle "Kestirici değişkenlerimizin etkilerini ortadan kaldırdıktan sonra geriye ne çok değişkenliğin kaldığını görmek bizi sarstı." Bu sıkı kontrollü ortamda bile durum gürültüsünü tam olarak hangi faktörlerin yönlendirdiği konusu gizemini koruyordu.

Araştırmacıların incelediği bütün değişkenler arasında, bir deneğin belli bir listedeki performansının en güçlü kestiricisi dış bir faktör değildi. Bir kelime listesindeki performans en iyi deneğin

ondan hemen önceki listede nasıl bir performans sergilediğiyle öngörülebiliyordu. Başarılı bir listenin arkasından yine nispeten başarılı bir liste, vasat bir listenin arkasından yine vasat bir liste gelmesi çok olasıydı. Performans listeden listeye rastgele değişmiyordu; her oturumda bariz bir dış neden olmadan, zaman içinde bir alçalıp bir yükseliyordu.

Bu bulgular Kahana ve arkadaşlarının kavramlarıyla hafıza performansının büyük ölçüde "hafıza işlevini yöneten iç kaynaklı sinirsel sürecin etkinliği" tarafından güdüldüğünü düşündürüyordu. Başka bir deyişle beynin etkinliğinin andan ana değişkenliğini sadece hava durumu ya da dikkat dağıtıcı bir müdahale gibi dış etkiler yönlendirmiyordu. Bu, beynimizin kendisinin işleme şeklinin bir özelliğiydi.

Beynin işlevindeki yerleşik değişkenliğin yargılarımızın kalitesini kontrol etmeyi umamayacağımız şekillerde etkilemesi çok olası. Beynin işlevindeki bu değişkenlik durum gürültüsünün bertaraf edilebileceğini düşünen herkesi duraksatmalı. Serbest atış çizgisindeki basketbol oyuncusuyla benzerlik başlangıçta göründüğü kadar basit değildi: Nasıl oyuncunun kasları hiçbir zaman aynı jesti üretmiyorsa, nöronlarımız da hiçbir zaman aynı şekilde işlemez. Zihnimiz bir ölçüm aracıysa, hiçbir zaman kusursuz bir ölçüm aracı olmayacaktır.

Bununla birlikte kontrol edilebilir lüzumsuz etkileri kontrol etmek için çabalayabiliriz. Bölüm 8'de göreceğimiz gibi bunu yapmak, yargılara grup hâlinde varıldığında özellikle önemlidir.

Durum Gürültüsünden Bahsederken

"Yargı bir serbest atış gibidir; aynısını tekrarlamak için ne kadar çabalarsak çabalayalım; hiçbir zaman aynısı olmaz."

"Yargınız ruh hâlinize, biraz önce değerlendirdiğiniz vakalara ve hatta hava durumuna bağlıdır. Her seferinde aynı insan olmazsınız."

"Geçen hafta olduğunuz kişi olmasanız da geçen haftaki "sizden" bugün bir başkasından olduğunuz kadar farklı olmazsınız. Durum gürültüsü sistem gürültüsünün en büyük kaynağı değildir."

BÖLÜM 8

Gruplar Gürültüyü Nasıl Artırır?

Bireysel yargıda gürültü zaten yeterince kötüdür. Ama grup hâlinde karar almak, soruna bir katman daha ekler. Gruplar, kısmen alakasız olması gereken faktörlere bağlı olarak çok farklı yönlere gidebilirler. Önce kim konuşacak, en son kim konuşacak, kim öz güvenle konuşuyor, kim siyah giymiş, kim kimin yanında oturuyor, kim gülümsüyor, kimin kaşları çatık ya da kim doğru anda doğru jestler yapıyor, bütün bunlar ve daha pek çoğu sonuçları etkiler. Her gün, konu ister işe alım olsun, ister terfi, bir ofisin kapatılması, iletişim stratejileri, çevresel düzenlemeler, ulusal güvenlik, üniversite kabulleri ya da piyasaya yeni ürün sürmek, benzer gruplar çok farklı kararlar alıyorlar.

Bir önceki bölümde çok sayıda bireyin yargılarını bir araya toplamanın gürültüyü azalttığını belirttiğimiz için, bu noktayı vurgulamak tuhaf görünebilir. Ancak grup dinamikleri nedeniyle, gruplar da gürültü ekleyebilirler. Ortalama yargısı doğru cevaba yakın olan "bilge kalabalıklar" olduğu gibi zorbaların peşine takılan, piyasa balonlarını körükleyen, sihre inanan ya da ortak bir yanılsamanın etkisi altında olan kalabalıklar da var. Küçük ve önemsiz farklılıklar bir grubu katı bir evet ve özünde birbirinin

aynısı olan bir grubu empatik bir hayır tavrına yönlendirebilir. Grup üyeleri arasındaki dinamikler nedeniyle –ki biz burada buna vurgu yapacağız– gürültü seviyesi yüksek olabilir. Bu önerme, benzer gruplar içindeki gürültüden bahsederken de, önemli bir konudaki sağlam yargısı sadece bir olasılıklar bulutu içindeki olasılıklardan biri gibi görülmesi gereken tek bir grup içinde geçerlidir.

Müzikte Gürültü

Kanıt olarak, alakasız gibi görünebilecek bir yerden başlayacağız ve Matthew Salganik ile çalışma arkadaşlarının müzik indirme üzerine yaptıkları geniş ölçekli bir çalışmadan bahsedeceğiz. Çalışma tasarlanırken, deney sahipleri binlerce kişiden oluşan bir kontrol grubu (orta düzeyde popüler bir internet sitesinin ziyaretçileri) yarattılar. Kontrol grubunun üyeleri yeni grupların yetmiş iki şarkısından bir ya da daha fazlasını dinleyebiliyor ve indirebiliyorlardı. Şarkılara canlı isimler verilmişti: "Bir Portakal Kabuğu İçinde Tutsak", "Kemirme", "Göz Bandı", "Beyzbol Warlock v1" ve "Pembe Saldırı" gibi. (İsimlerden bazıları kulağa bizim buradaki ilgi alanımızla doğrudan ilişkili geliyor: "En İyi Hatalar", "Ben Hatayım", "Cevabı Aşan İnanç", "Hayatın Gizemi", "Bana Şans Dile" ve "Ormanın Dışında".)

Kontrol grubunda katılımcılara başkalarının ne dediği ya da yaptığı konusunda hiçbir şey söylenmedi. Hangi şarkıları sevdikleri ve indirmek istedikleri konusunda bağımsız kararlar verebileceklerdi. Ancak Salganik ve arkadaşları, diğer internet sitesi ziyaretçilerinin rastgele dağıtıldığı sekiz grup daha yarattı. Bu grupların üyeleri için tek bir istisnayla her şey aynıydı: İnsanlar her şarkıyı kendi gruplarından, daha önce kaç kişinin indirdiğini görebilecekti. Örneğin "En İyi Hatalar" bir grupta çok popülerse ya da kimse indirmediyse, grubun üyeleri bunu görebilecekti.

Çeşitli gruplar herhangi bir önemli boyutta farklılık sergilemediği için, çalışma, geçmişi sekiz kez işletiyordu. Sonunda en iyi şarkıların her zaman en tepeye çıkacağını ve kötü olanların her zaman dibe batacağını tahmin edebilirsiniz. Böyle olsaydı, çeşitli gruplar sonunda aynı ya da en azından benzer puanlamalara ulaşırdı. Grup içinde gürültü olmazdı. Zaten Salganik ve arkadaşlarının cevabını aradıkları soru tam olarak buydu. Gürültü etmenlerinden özellikle birini test ediyorlardı: *sosyal etki.*

Kilit bulgu, grup sıralamalarının çılgınca farklılık sergilediği oldu. Farklı gruplar arasında çok fazla gürültü vardı. Bir grupta "En İyi Hatalar" önemli bir başarıya imza atarken, "Ben Hatayım" çuvallayabiliyordu. Bir başka grupta "Ben Hatayım" iyi iş çıkarırken, "En İyi Hatalar"ın durumu feci olabiliyordu. Bir şarkı başlangıçta popülerlik yakalarsa çok iyi sonuç alıyordu. Ama bu olmadığında, sonuç çok farklı olabiliyordu.

Gerçekten de en kötü şarkılar (kontrol grubunun belirlediği hâliyle) hiçbir zaman zirveye çıkamıyordu ve en iyi şarkılar da asla dibi boylamıyordu. Ancak bunun dışında hemen her şey mümkündü. Yazarların vurguladığı gibi "Sosyal etki şartında başarı seviyesi bağımsız şartta olduğunda daha öngörülemezdi." Kısaca sosyal etki gruplar arasında önemli gürültü yaratıyordu. Şöyle bir düşündüğünüzde bir şarkı lehine ya da aleyhine yargılarının, şarkının erken dönemde popülerlik kazanıp kazanmamasına bağlı olarak değişmesinin çok kolay olması anlamında, grupların da kendi içinde gürültülü olduğunu görebilirsiniz.

Salganik ve çalışma arkadaşlarının daha sonra gösterdiği gibi grup sonuçlarının manipüle edilmesi çok kolaydır çünkü popülerlik kendini pekiştirir. Bunu izleyen kurnazca sayılabilecek bir deneyde, kontrol grubundaki sıralamaları tersine çevirdiler (başka bir deyişle şarkıların popülerliği konusunda yalan söylediler). Yani insanlar en az popüler şarkıları en popüler olarak görecekti (bunun tersi de doğruydu). Araştırmacılar daha sonra internet sitesi

ziyaretçilerinin ne yapacağını test etti. Sonuçta en az popüler şarkıların çoğu popülerlik kazanırken, en popüler şarkılar düşüşe geçti. Çok büyük grupların içinde popülerlik ve popüler olmama durumu, araştırmacıların insanları hangi şarkıların popüler olduğu konusunda yanlış yönlendirdiği zaman bile aynı sonucu üretti. Tek istisna, kontrol grubunun en popüler şarkısının zaman içinde popülerliğini artırması oldu. Yani sıralamayla oynanması en iyi şarkıyı dipte tutmaya yetmedi. Ancak, sıralamanın tersine çevrilmesi, ekseriyetle nihai sıralamanın belirlenmesine yardımcı oldu.

Bu çalışmaların genel olarak grup yargısına yansımasını görmek kolay olmalı. Varsayalım on kişiden oluşan küçük bir grup yeni ve cesur bir girişimi benimseyip benimsememek konusunda karara varmaya çalışıyor olsun. Önce bir ya da iki savunucu konuşursa, odanın tamamını tercih ettikleri yöne kaydırabilirler. Aynı şey şüphecilerin önce konuşması durumunda da geçerlidir. En azından insanlar birbirlerinden etkilendiklerinde böyle olur ve genelde insanlar birbirlerinden etkilenirler. Bu nedenle birbirine benzer gruplar, ilk kimin konuştuğuna ve ilk indirilen şarkıların dengi bir girişimde bulunduğuna bağlı olarak çok farklı yargılara varabilirler. "En İyi Hatalar"ın ve "Ben Hatayım"ın popülerliklerinin benzerleri her tür profesyonel yargıda bulunabilir. Gruplar bu tür şarkıların popülerlik sıralamasının bir benzerini duymazlarsa –cesurca girişim için yüksek sesli heyecan ifadesi gibi– girişim, sırf savunucuları görüşlerini dile getirmediği için hiçbir yere varmayabilir.

Müzik İndirmenin Ötesi

Şüpheci bir insansanız, müzik indirme örneğinin benzersiz ya da en azından kendine özgü olduğunu ve bize diğer grupların vardığı yargılar hakkında çok az şey anlattıklarını düşünüyor olabilirsiniz. Ancak daha pek çok alanda benzer gözlemler yapıldı.

Örneğin Birleşik Krallık'ta referandum tekliflerinin popülerliğini ele alalım. Bir referanduma destek verip vermeyeceklerine karar verirken, insanların elbette her şey göz önüne alındığında iyi bir fikir olup olmadığını değerlendirmesi gerekir. Örüntüler Salganik ve arkadaşlarının gözlemlediklerine benzer: Başlangıçtaki bir popülerlik çıkışı kendi kendini pekiştiricidir ve bir öneri ilk gününde az destek görürse, sonu parlak olmaz. Müzikte olduğu gibi politikada da sosyal etkiler ve özellikle insanların diğer insanların o fikri cazip ya da itici bulduğunu görüp görmemeleri çok şeyi değiştirir.

Cornell Üniversitesi'nden sosyolog Michael Macy ve çalışma arkadaşları, doğrudan müzik indirme deneyinden yola çıkarak, diğer insanların görünebilir görüşlerinin tanımlanabilir politik görüşleri Demokratlar arasında popüler ve Cumhuriyetçiler arasında itici –ya da tam tersi– kılıp kılamayacağını sorguladılar. Kısa cevap evet oldu. Demokratlar çevrim içi bir grupta belli bir görüşün başlangıç itibarıyla Demokratlar arasında popülerlik kazandığını gördükleri zaman, o görüşü destekliyor ve nihayetinde ilgili gruptaki Demokratların çoğu o görüşün lehine dönebiliyordu. Ancak farklı bir çevrim içi gruptaki Demokratlar aynı görüşün başlangıç itibarıyla Cumhuriyetçiler arasında popülerlik kazandığını gördüklerinde, onu derhâl reddediyor ve nihayetinde ilgili gruptaki Demokratların çoğu görüşün aleyhine dönüyordu. Cumhuriyetçiler de benzer bir davranış içindeydiler. Kısaca politik görüşler de nihai yazgılarının başlangıçtaki popülerlik düzeyine bağlı olabilmesi açısından, şarkılara benziyorlar. Araştırmacıların kendi ifadesiyle "başlangıçtaki az sayıdaki girişimci grubundaki şans değişkeni" büyük topluluklarda dengeyi değiştirme ve hem Cumhuriyetçilere hem Demokratlara birbiriyle hiç alakası olmayan bir görüş öbeğini benimsetme etkisine sahip olabilir.

Ya da doğrudan genel olarak grup kararlarına etki eden bir soruyu ele alalım: İnsanlar internet sitelerinde yer alan yorumları nasıl yargılar? Kudüs İbrani Üniversitesi'nden profesör Lev Much-

nick ve çalışma arkadaşları, farklı hikâyeler sergileyen ve insanların oylamayla yukarı ya da aşağı taşınabilen yorumlar bırakmasına imkân sağlayan bir internet sitesinde bir deney gerçekleştirdiler. Araştırmacılar otomatik ve yapay olarak, hikâyelere bırakılan belli yorumlara, ilk oyu olumlu olarak verdiler. Yüzlerce ya da binlerce ziyaretçi ve puanlamadan sonra, bir yoruma verilen ilk oyun önemli olmasının mümkün olmayacağını düşünebilirsiniz. Makul bir düşünce ama yanlış. İlk olumlu oyu gördükten sonra (ve tamamen yapay olduğunu hatırlatalım), bir sonraki ziyaretçinin de olumlu oy verme ihtimali %32 artıyordu.

Dikkat çekici bir şekilde, bu etki çok uzun süre dayanıyordu. Beş ayın sonunda, ilk verilen tek bir olumlu oy yorumların ortalama puanını %25 artırıyordu. Başlangıçta verilen tek bir olumlu oyun etkisi bir gürültü tarifidir. O oyun verilme nedeni ne olursa olsun, toplam popülerlikte büyük ölçekli bir kaymaya yol açabilir.

Bu çalışma grupların nasıl yön değiştirdiği ve neden gürültülü olduğu konusunda (yine benzer grupların çok farklı yargılara varabilmesi ve tekil grupların yargısının bir olasılık bulutu içinden seçilmiş herhangi bir tanesi olabilmesi anlamında) bir ipucu veriyor. Üyeler onay vererek, nötr kalarak ya da karşıt görüş bildirerek, kendilerini başlangıçta verilen olumlu bir oyun (ya da olumsuz bir oyun) işlevsel dengini sunarken bulabilirler. Bir grup üyesi hemen onay verirse, diğer üyelerin de aynı şeyi yapmak için nedeni olur. Gruplar belli ürünlere, insanlara, hareketlere ya da fikirlere doğru hareket ettiklerinde, bunun nedeni söz konusu şeylerin asli değerleri değil, verdiğimiz örnekteki ilk oyların işlev açısından dengi olan bir hareket olabilir. Elbette Muchnik'in çalışması çok büyük gruplar üzerinde yapıldı. Ancak aynı şey küçük gruplarda, hatta belki daha çarpıcı şekilde yaşanabilir, çünkü genellikle başlangıçtaki ilk oyun bir plan, ürün ya da karar lehine olmasının diğerleri üstünde büyük etkisi olur.

Bununla ilgili bir nokta daha var. Kalabalıkların bilgeliğinden bahsetmiştik: Büyük bir insan grubunu alır ve onlara bir soru sorarsanız, ortalama cevabın hedefe yakın olması ihtimali yüksektir. Yargıları bir araya toplamak gürültüyü, dolayısıyla hatayı azaltmanın harika bir yolu olabilir. Peki, insanlar birbirini dinlediği zaman ne olur? Birbirlerini dinlemelerinin faydalı olacağını düşünebilirsiniz. Sonuçta insanlar birbirlerinden bir şeyler öğrenip neyin doğru olduğunu çözebilirler. İnsanların bildiklerini paylaştığı olumlu şartlarda, müzakere eden gruplar iyi iş çıkarabilir. Ancak bağımsızlık kalabalıkların bilgeliği için ön şarttır. İnsanlar kendi yargılarına varmak yerine başka insanların ne düşündüğüne göre hareket ederlerse, kalabalıklar o kadar da bilge olmayabilir.

Araştırmalar tam olarak bu sorunu açığa çıkardı. Basit tahmin görevlerinde –bir şehirde işlenen suç sayısı, belirli dönemlerde nüfus artışı ve ülkeler arasındaki bir sınırın uzunluğu gibi– kalabalıklar görüşlerini bağımsız olarak bildirdikleri sürece, gerçekten bilgeydiler. Öte yandan diğer insanların tahminlerini öğrenirlerse –örneğin on iki kişilik bir grubun ortalama tahmini– kalabalık daha kötü iş çıkarıyordu. Yazarların ifadesiyle sosyal etkiler "toplam hatayı azaltmadan grup çeşitliliği"ni azalttıkları için sorun oluşturuyorlardı. Burada asıl ironi şu: Layıkıyla bir araya toplanan çok sayıda bağımsız görüş çarpıcı bir doğruluk yakalayabilirken, küçük bir sosyal etki bile kalabalıkların bilgeliğini baltalayan bir tür güdüme neden olabiliyor.

Şelaleler

Anlattığımız çalışmaların bazıları *bilgilendirici şelaleler* içeriyor. Bu tür şelaleler her yerdedir. İş dünyasında, hükûmette ve diğer yerlerde benzer grupların neden birçok farklı yöne gidebildiğinin ve küçük değişikliklerin neden bu kadar farklı sonuçlar ve dolayısıyla gürültü yarattığının izah edilmesine yardım ederler. Bizler

geçmişi gerçekte yaşandığı hâliyle görebiliyoruz ama pek çok grup ve grup kararında, olasılık bulutlarından sadece biri gerçekleşir.

Bilgilendirici şelalelerin nasıl işlediğini anlamak için, büyük bir ofiste, önemli bir pozisyon için kimi işe alacaklarına karar veren on kişi olduğunu düşünün. Öne çıkan üç aday var: Thomas, Sam ve Julie. Grup üyelerinin görüşlerini peş peşe ilan ettiklerini varsayın. Her biri, makul bir şekilde, diğerlerinin yargılarını dinliyor. İlk sözü Arthur alıyor. En iyi tercihin Thomas olacağını söylüyor. Barbara artık Arthur'un yargısını biliyor; o da Thomas konusunda hevesliyse elbette Arthur'un görüşüne uymalı. Ancak diyelim Barbara henüz en iyi adayın kim olduğu konusunda emin değil. Arthur'a güveniyorsa, onu onaylayabilir: Evet, Thomas en iyisi. Arthur'a yeterince güvendiği için yargısını da destekliyor.

Şimdi üçüncü kişiye, Charles'a dönelim. Arthur da Barbara da işe Thomas'ı almak istediklerini söylediler ama sınırlı olduğunu bildiği bilgisine dayanarak Charles'ın görüşü Thomas'ın bu görev için doğru insan olmadığı ve en iyi adayın Julie olduğu yönünde olsun. Charles bu görüşe sahip olmasına rağmen, bildiği şeyi yok sayarak Arthur ve Barbara'nın peşinden gidebilir. Bunu yaparsa, nedeni Charles'ın korkak olması değil, saygılı bir dinleyici olmasıdır. Hem Arthur hem Barbara'nın heyecanları için yeterli kanıtlarının olduğunu düşünmüş olabilir.

David kendi bilgisinin ondan önce gelen kişilerden gerçekten daha iyi olduğunu düşünmediği sürece, onların görüşünün peşinden gitmelidir ve öyle de yapacaktır. Bunu yaparsa bir şelaleye girmiş olur. Doğru, Arthur, Barbara ve Charles'ın yanıldıklarını düşünmek için güçlü dayanağı olması hâlinde, karşı koyacaktır. Ancak böyle bir dayanağı yoksa, onlara uyması muhtemeldir.

Charles ve David, Thomas (ya da diğer adaylar) hakkında Arthur ve Barbara'nın farkında olmadığı bilgi ve içgörüye sahip olabilirler. Bu özel bilgi paylaşılsaydı, Arthur ya da Barbara'nın fikrini değiştirebilirdi. İlk konuşan Charles ya da David olsaydı,

sadece adaylar hakkındaki görüşlerini ifade etmekle kalmayabilir ve diğer katılımcıları etkileyebilecek bilgi katkısında bulunabilirlerdi. Ancak en son konuştukları için, kendilerine ait bilgi pekâlâ onlarda kalabilir.

Şimdi Erica, Frank ve George'un görüşlerini ifade etmelerinin beklendiğini varsayın. Arthur, Barbara, Charles ve David daha önce Thomas'ın en iyisi olduğunu söyledilerse Erica, Frank ve George da, başka bir seçeneğin daha iyi olabileceğini düşünmek için iyi bir nedenleri olsa bile onlarla aynı şeyi söyleyebilirler. Elbette açıkça yanlış olsa, gittikçe büyüyen konsensüse itiraz edebilirler. Ama ya karar net değilse? Bu örnekteki hassas nokta, Arthur'un başlangıçtaki yargısının birçok insanın, grubun –bazılarının aslında bir görüşü olmamasına ve diğerlerinin en iyi tercih olmadığını düşünmesine rağmen– oy birliğiyle Thomas'ta karar kılmasına yol açan bir şelaleye katılmasına neden olan süreci başlatması.

Elbette bu son derece yapay bir örnek. Ancak her tür grupta buna benzer şeyler her zaman olur. İnsanlar başkalarından bir şeyler öğrenir ve ilk konuşmacılar bir şeyden hoşlanır ya da bir şeyi yapmak ister gibi görünürse, diğerleri rıza gösterebilir. En azından, onlara güvenmemek için bir nedenleri ve yanıldıklarını düşünmek için iyi bir sebepleri olmadığında böyle olur.

Bizim amaçlarımız açısından buradaki en önemli kısım, bilgilendirici şelalelerin gruplar içinde gürültüyü olası hatta mümkün kılması. Verdiğimiz örnekte Arthur ilk sözü aldı ve oyunu Thomas'tan yana kullandı. Ama ilk konuşanın Barbara olduğunu ve Sam'den yana olduğunu varsayın. Ya da Arthur'un biraz farklı hissettiğini ve Julie'yi tercih ettiğini. Makul varsayımlarda grup, daha iyi oldukları için değil, şelale bu şekilde işleyeceği için Sam ya da Julie'ye yönelebilirdi. Müzik indirme deneyinin (ve kuzenlerinin) merkez bulgusu budur.

İnsanların bilgilendirici şelalelere katılmasının illa mantıksız olmadığını belirtelim. İnsanlar işe kimi alacakları konusunda emin

değillerse, diğerlerine uymak akıllıca olabilir. Aynı görüşü paylaşan insanların sayısı arttıkça, onlara uymak daha da akıllıca olur. Bununla birlikte burada iki sorun bulunuyor. İlk olarak insanlar kalabalıktaki insanların çoğunun da bir şelalede olabileceği ve kendi başlarına bağımsız yargılara varmadıkları olasılığını yok sayma eğilimindedir. Bir sonucu benimseyen üç, on ya da yirmi insan gördüğümüzde, her birinin kendilerinden önce gelenleri ne dereceye kadar takip edeceklerini hafife alabiliriz. Aslında sadece birkaç kişinin başlangıçtaki görüşlerini yansıtsa da ortak görüşlerinin toplu bilgeliklerini yansıttığını düşünebiliriz. İkinci olarak bilgilendirici şelaleler insan gruplarını gerçek anlamda korkunç yönlere sürükleyebilir. Sonuçta Arthur, Thomas konusunda yanılmış da olabilir.

Elbette grup üyelerinin birbirlerinden etkilenmesinin tek nedeni bilgi değil. Sosyal baskı da önemlidir. Bir şirkette ya da hükûmette insanlar sevimsiz, gaddar, anlayışsız ya da aptal görünmemek için kendilerini susturabilirler. Takım oyuncusu olmak isterler. Başkalarının görüşlerini ve eylemlerini bu yüzden takip ederler. İnsanlar neyin doğru ve muhtemelen doğru olduğunu bildiklerini düşünürler ama yine de grupla iyi geçinmek için, grubun belirgin konsensüsüne ya da ilk konuşmacıların görüşlerine uyarlar.

Ufak sapmalarla, az önce anlattığımız bu işe alım hikâyesi insanlar birbirlerinden Thomas'ın artıları hakkında bir şeyler öğrendiği için değil, geçimsiz ya da aptal görünmek istemedikleri için de aynı şekilde ilerleyebilir. Arthur'un en başta Thomas hakkında olumlu yargı belirtmesi bir çoğunluğa uyma etkisi başlatarak, nihayetinde, sırf diğer herkes Thomas'tan yana olduğu için Erica, Frank ya da George üzerinde güçlü bir sosyal baskıya sebep olabilir. Ve bilgilendirici şelalelerde olduğu gibi sosyal baskı şelalelerinde de insanlar onlardan önce söz alan kişilerin kanaatlerini abartabilirler. İnsanlar Thomas'ı destekliyorsa bunu gerçekten Thomas'ı yeğledikleri için değil, ilk konuşmacılardan ya da güçlülerden biri

onu desteklediği için yapıyor olabilirler. Bununla birlikte grup üyeleri sonunda kendilerini, konsensüse kendi seslerini eklerken ve böylece sosyal baskının seviyesini artırırken bulabilirler. Bu, şirketler ve hükûmet büroları için çok tanıdık bir fenomendir ve aslında yanlış olan bir yargı konusunda güven yaratılmasının ve oy birliğiyle verilen bir desteğin yolunu açabilir.

Gruplar arasında, sosyal etkiler de gürültü yaratır. Bir kişi bir toplantıyı şirketin yönetiminde önemli bir değişikliğin lehinde görüş bildirerek başlatırsa, o kişi bir grubun o değişikliği oy birliğiyle desteklemesine yol açacak bir tartışmanın fitilini ateşlemiş olabilir. Varılan anlaşma inancın değil, sosyal baskıların ürünü olabilir. Toplantıyı bir başkası farklı bir görüşü işaret ederek açsaydı ya da ilk konuşan kişi sessiz kalmaya karar verseydi, tartışma, yine aynı sebeple tamamen başka bir yöne yönelebilirdi. Çok benzer gruplar sosyal baskılar yüzünden kendilerini birbiriyle çok uyumsuz yerlerde bulabilirler.

Grup Kutuplaşması

ABD ve daha birçok ülkede kamu davaları (ve diğer pek çok hukuki dava) jüri eşliğinde görülür. İnsan, müzakerelerinin sonucunda, jürilerin bu müzakere üstüne kurulu toplulukları oluşturan bireylerden daha akıllıca kararlar almasını umuyor. Ancak jürilerin incelenmesi, yine bir gürültü kaynağı olan ayrı bir sosyal etki türünü, *grup kutuplaşmasını* karşımıza getiriyor. Temel düşünce insanların birbirleriyle konuştukları zaman, sıklıkla başlangıçtaki eğilimleriyle aynı doğrultuda, daha uç bir noktaya varmalarıdır. Örneğin yedi kişilik bir grupta insanların çoğu Paris'te bir ofis açmanın bayağı iyi bir fikir olacağını düşünme eğilimindeyse, tartışmanın ardından, grubun ofisi açmanın muhteşem bir fikir olacağı sonucuna varması olasıdır. İçsel tartışmalar sık sık daha fazla güven, daha fazla birlik ve sıklıkla hevesin artması şeklinde

daha uç bir duruşa yol açar. Gelin görün ki grup kutuplaşması sadece jürilerde yaşanmaz; profesyonel kararlar alan takımlar da sık sık kutuplaşır.

Bir dizi deneyde, ürün sorumluluğu vakalarında cezai tazminattan yana olan jürilerin kararlarını inceledik. Her jürinin kararı, şirketi kusurlu davranış için cezalandırmanın yanı sıra başkaları için de caydırıcı olması hedeflenen bir miktardı. (Bölüm 15'te bu çalışmalara dönecek ve onları detaylı olarak anlatacağız.) Buradaki amaçlarımız gereği, gerçek dünyadaki müzakere eden jürileri ve "istatistiki" jürileri karşılaştıran bir deney düşünün. İlk olarak, çalışmamızdaki 899 katılımcıya vaka özetleri sunduk ve her vaka için, öfkelerini ve ceza niyetlerini ifade etmek için 7 derecelik bir ölçek ve (varsa) para cezası için de bir dolar ölçeği kullanarak kendi bağımsız yargılarını belirlemelerini istedik. Ardından, bilgisayar yardımıyla bu bireysel cevapları milyonlarca istatistiki jüri, yani altı kişilik sanal gruplar (rastgele bir araya getirilmiş) yaratmak için kullandık. Her istatistiki jüride altı bireysel yargının medyanını nihai hüküm olarak kabul ettik.

Kısaca bu istatistiki jürilerin yargılarının çok daha tutarlı olduğunu gördük. Gürültü önemli ölçüde azalmıştı. Alçak gürültü istatistiki birleştirmenin mekanik bir etkisiydi: Bağımsız, bireysel yargılarda var olan gürültü, ortalamalarının alınmasıyla her zaman azalır.

Öte yandan gerçek dünya jürileri, istatistiki jüriler değiller; bir araya gelip vaka hakkındaki görüşlerini tartışırlar. Makul bir şekilde, müzakere eden jürilerin medyan üyelerinin yargısına ulaşma eğiliminde olup olmadığını merak ediyor olabilirsiniz. Bunu öğrenmek için, ilk deneyin arkasından, jüriye seçilebilir özellikteki üç bin vatandaş ve beş yüzden fazla altı kişilik jüri içeren bir deney daha gerçekleştirdik.

Sonuçlar çok açıktı. Aynı vakaya bakarken, müzakereci jüriler istatistiki jürilerden çok daha gürültülüydüler: Bu, sosyal etki

gürültüsünün açık bir yansımasıydı. Müzakere gürültüyü artırıcı etkiye sahipti.

İlgi çekici bir bulgu daha vardı. Altı kişilik bir grubun medyan üyesi sadece biraz öfkelendiği ve daha yumuşak bir cezayı tercih ettiğinde, müzakere eden jürinin hükmü tipik olarak daha da yumuşak oluyordu. Aksine, altı kişilik grubun medyan üyesi daha öfkeli olup daha katı bir ceza tercihini ifade ettiğinde, müzakere eden jüri tipik olarak daha öfkeli ve daha katı bir tavra giriyordu. Bu öfke para cezası olarak ifade edildiğinde ise sistematik eğilim, jürinin medyan üyesinin karar kıldığı cezadan daha yüksek bir para cezası oluyordu. Hatta jürilerin %27'si en katı üyeleri kadar, hatta ondan da yüksek bir ceza seçiyordu. Müzakereci jüriler istatistiki jürilerden sadece daha gürültülü değildiler, aynı zamanda onları oluşturan bireylerin görüşlerini de kuvvetlendiriyorlardı.

Grup kutuplaşmasının temel bulgusunu hatırlayın: İnsanlar birbirleriyle konuştuktan sonra, tipik olarak başlangıçtaki eğilimleriyle aynı doğrultuda daha uç bir noktaya ulaşırlar. Deneyimiz bu etkiyi resmediyor. Müzakereci jüriler medyan üye merhametli olduğunda daha merhametli ve medyan üye katı olduğunda daha katı tutumlara kayma yaşadılar. Benzer şekilde, parasal cezalar vermeye meyilli jüriler sonunda medyan üyelerinin tercih ettiğinden daha katı cezalar verdiler.

Grup kutuplaşmasının açıklamaları şelale etkilerinin açıklamalarına benzer. Bilgi büyük rol oynar. İnsanların çoğu ağır bir cezadan yanaysa, grup ağır ceza lehine çok fazla argüman, diğer istikamette ise çok daha az argüman duyacaktır. Grup üyeleri birbirlerini dinliyorlarsa, baskın eğilim yönüne kayarak grubu daha birleşik, daha güvenli ve daha ekstrem kılarlar. İnsanlar grup içindeki itibarlarını önemsiyorlarsa yine baskın eğilim yönüne kaymaları da kutuplaşma üretir.

Grup kutuplaşması elbette hata üretebilir. Sıklıkla üretir de. Ancak bizim buradaki başlıca odağımız değişkenlik. Gördüğümüz

gibi, yargıların birleştirilmesi gürültüyü azaltır ve bu açıdan, ne kadar çok yargı olursa o kadar iyidir. İstatistiki jürilerin bireysel jüri üyelerinden daha az gürültülü olması bu yüzdendir. Aynı zamanda, müzakereci jürilerin istatistiki jürilerden daha gürültülü olduğunu da öğrendik. Benzer özelliklerdeki gruplar sonunda farklılaştıklarında, sıklıkla nedeni grup kutuplaşmasıdır. Ve bunun sonucunda ortaya çıkan gürültü çok yüksek olabilir.

İş dünyasında, hükûmette ve diğer yerlerde, şelaleler ve kutuplaşma aynı soruna bakan gruplar arasında uyuşmazlıklara yol açabilir. Bireysel yargıların ne kadar gürültülü olduğunu keşfettiğimiz bu noktada, sonuçların birkaç bireyin (ilk konuşan ve en çok etkiye sahip olanlar) yargılarına bağlı olma potansiyeli bizi özellikle endişelendirmeli. Yatay gürültü ve örüntü gürültüsünün grup üyelerinin yargıları arasında olması gerekenden (ve normalde beklediğimizden) daha büyük farklar yarattığını gördük. Ayrıca durum gürültüsünün –yorgunluk, ruh hâli, karşılaştırma noktaları gibi– konuşan ilk kişinin yargısını etkileyebileceğini de biliyoruz. Grup dinamikleri gürültüyü şiddetlendirebilir. Sonuç olarak, müzakereci gruplar bireysel yargıların ortalamasını alan istatistiki gruplardan daha gürültülü olma eğilimindedir.

İş dünyasındaki ve hükûmetlerdeki çoğu önemli karar bir tür müzakere sürecinin sonunda alındığı için, bu riske karşı uyanık olmak özellikle önemlidir. Örgütler ve liderleri bireysel üyelerinin yargılarındaki gürültüyü kontrol etmek için adımlar atmalılar. Ayrıca müzakereci grupları gürültüyü şiddetlendirmek yerine, azaltacak şekilde idare etmeliler. Önereceğimiz gürültüyü azaltma stratejileri bu hedefe ulaşmayı amaçlıyor.

Grup Kararlarından Bahsederken

"Her şey başlangıçtaki popülerliğe dayanıyor gibi görünüyor. Piyasaya yeni sürdüğümüz her şeyin, muhteşem bir ilk hafta geçirmesini sağlamak için çok çalışmalıyız."

"Her zaman şüphe duyduğum gibi politika ve ekonomiyle ilgili düşünceler, film yıldızlarına benziyor. İnsanlar başka insanların onları beğendiğini düşünürse, bu tür fikirler çok yol katedebiliyor."

"Takımım bir araya toplandığında, toplantıdan emin, tamamen aynı görüşte ve seçtiğimiz eylem planına tamamen adanmış bir hâlde çıkmamız beni her zaman endişelendirmiştir. İç süreçlerimizde çok iyi gitmeyen bir şeyler olsa gerek!"

KISIM 3

Öngörücü Yargılarda Gürültü

Pek çok yargı öngörüdür ve doğrulanabilir öngörüler değerlendirilebildiği için, onları inceleyerek gürültü ve ön yargı hakkında çok şey öğrenebiliriz. Kitabın bu kısmında, öngörücü yargılara odaklanacağız.

Bölüm 9 profesyonellerin, makinelerin ve basit kuralların öngörülerinin isabetliliğini karşılaştırıyor. Bu yarışta profesyonellerin ancak üçüncü sırada geldiği sonucuna ulaşmamız sizi şaşırtmayacaktır. Bölüm 10'da bu sonucun nedenlerini araştıracak ve gürültünün insan yargısının daha alt seviyede kalmasında önemli bir faktör olduğunu göstereceğiz.

Bu sonuçlara ulaşmak için öngörülerin kalitesini değerlendirmeliyiz. Bunu yapmak için de öngörü isabetliliği için bir ölçüye, şu soruyu bir şekilde cevaplamaya ihtiyacımız var: Öngörüler sonuçlarla *birlikte değişiyor* mu? Değişiyorsa ne kadar sık değişiyor? Örneğin İK departmanı yeni işe alınan çalışanların potansiyelini rutin olarak puanlarsa, çalışanların nasıl bir performans sergilediğini öğrenmek için birkaç yıl bekleyebilir ve potansiyel puanla-

rının performans değerlendirmeleriyle ne kadar yakın değiştiğini görebiliriz. Öngörüler, potansiyeli yüksek puanlanan çalışanların işe alındıkları zaman çalışmalarına karşılık da yüksek değerlendirmeler kazandığı ölçüde doğrudur.

Bu sezgiyi yansıtan bir ölçü, *yüzdelik uyumdur (YU)* ve daha belirli bir soruyu cevaplar: Rastgele bir çift çalışan seçtiğinizi varsayın. Potansiyel değerlendirmesinde daha yüksek puan alan çalışanın iş başında da daha iyi performans sergileme olasılığı nedir? Başlangıçtaki puanlamanın isabetliliği kusursuz olsaydı; YU %100 olurdu: Yani iki çalışanın potansiyele göre puanlaması, daha sonraki performans değerlendirmelerinin kusursuz bir öngörüsü olurdu. Öngörüler tamamen işe yaramaz olsaydı, uyum sadece tesadüfen gerçekleşirdi ve "daha yüksek potansiyelli" çalışanın daha iyi performans sergilemesi de bir o kadar olasılık dışı olurdu: Yani YU %50 olurdu. Kapsamlı olarak incelenen bu örneği Bölüm 9'da tartışacağız. Daha basit bir örnek vermek gerekirse yetişkin erkeklerde ayak uzunluğu ve boy için YU %71'dir. İki insana –önce kafalarına, sonra ayaklarına– baktığınızda daha uzun boylu olanın ayaklarının da daha büyük olma olasılığı %71'dir.

YU, ilişkili değişimin hızlı bir şekilde sezgisel ölçümüdür ve çok büyük bir avantajdır ama sosyal bilimcilerin kullandığı standart ölçüm değildir. Standart ölçüm, iki değişken arasında pozitif bağıntı olduğunda 0 ve 1 arasında değişen *korelasyon katsayısı*dır *(r)*. Yukarıdaki örnekte boy ve ayak ölçüsü arasındaki korelasyon yaklaşık 0,60'tır.

Korelasyon katsayısını düşünmenin pek çok yolu vardır. Bunlardan yeterince sezgisel olanı şudur: İki değişken arasındaki korelasyon, ortak belirleyicilerinin yüzdesidir. Örneğin bir özelliğin tamamen genetik olarak belirlendiğini varsayalım. O özellik için genlerinin %50'si ortak olan kardeşler arasında 0,50'lik, genlerinin %25'i ortak olan kuzenler arasında ise 0,25'lik bir korelasyon bulmayı bekleriz. Ayrıca boy ve ayak ölçüsü arasındaki 0,60'lık

korelasyonu, boyu belirleyen nedensel faktörlerin %60'ının ayakkabı ölçüsünü de belirlediğine işaret ettiği şeklinde okuyabiliriz.

Tanımladığımız iki ilişkili değişim ölçüsü birbirleriyle doğrudan ilgilidir. Tablo 1 korelasyon katsayısının farklı değerleri için YU'yu sunuyor. Bu kitabın geri kalanında, insanların ve modellerin performansını tartışırken, her zaman iki ölçümü birlikte sunacağız.

Tablo 1: *Korelasyon katsayısı ve Yüzdelik uyum (YU)*

Korelasyon katsayısı	*Yüzdelik uyum (YU)*
0,00	%50
0,10	%53
0,20	%56
0,30	%60
0,40	%63
0,60	%71
0,80	%79
1,00	%100

Bölüm 11'de öngörü isabetliliğinin önemli bir sınırını, geleceğin bağlı olduğu pek çok şeyin bilinemez olması nedeniyle çoğu yargıya *objektif bilgisizlik* olarak adlandırdığımız bir durum içinde varıldığını konuşacağız. Çarpıcı bir şekilde, çoğu zaman bu sınırlandırmadan bihaber kalmayı ve güvenle (ya da daha doğrusu aşırı bir güvenle) öngörülerde bulunmayı başarırız. Son olarak Bölüm 12'de objektif bilgisizliğin sadece olayları öngörme becerimizi değil, onları anlama kapasitemizi bile etkilediğini göstereceğiz. Bu, gürültünün neden görünmez kalma eğiliminde olduğu bilmecesinin önemli bir parçasıdır.

BÖLÜM 9

Yargılar ve Modeller

Pek çok insan, insanların –hem kendilerinin hem başkalarının– işteki gelecek performansları üstüne tahminlerde bulunmakla ilgilenir. Bu nedenle, performans tahmini öngörücü yargının faydalı bir örneğidir. Misal büyük bir şirketteki iki üst düzey yöneticiyi düşünün. Monica ve Nathalie işe alındıkları zaman uzman bir danışmanlık şirketi tarafından değerlendirildiler ve liderlik, iletişim, insan ilişkileri, işle ilgili teknik beceriler ve bir sonraki pozisyon için motivasyon başlıklarında 1 ile 10 arası bir ölçekte puanlamalar aldılar (Tablo 2). Sizin göreviniz işe alınmalarından 2 sene sonrası için performans değerlendirmelerini, yine 1 ile 10 arası bir ölçek kullanarak öngörmek.

Tablo 2: *Üst düzey yönetici pozisyonu için iki aday*

	Liderlik	*İletişim*	*İnsan ilişkileri*	*Teknik beceriler*	*Motivasyon*	*Öngörünüz*
Monica	4	6	4	8	8	
Nathalie	8	10	6	7	6	

Çoğu insan bu tip bir sorunla karşılaştığında, basitçe her satıra dikkatle bakar ve bazen puanların ortalamasını zihnen hesaplayarak hızlı bir yargı üretir. Siz de bunu yaptıysanız, muhtemelen, Nathalie'nin daha güçlü aday olduğu ve Monica ile arasındaki farkın 1 ya da 2 puan olduğu sonucuna varmışsınızdır.

Yargı mı Formül mü?

Bu soruna takındığınız gayriresmî yaklaşım, *klinik yargı* olarak bilinir. Bilgiyi değerlendirir, belki hızlı bir hesaba girer, sezginize danışır ve bir yargıya ulaşırsınız. Aslında klinik yargı, bu kitapta basitçe yargı olarak tanımladığımız süreçtir.

Şimdi öngörü görevini bir deneyde katılımcı olarak gerçekleştirdiğinizi düşünün. Monica ve Nathalie birkaç yıl önce işe alınmış ve beş farklı boyutta puanlanmış yüzlerce müdürün olduğu bir veri tabanından alındılar. Siz bu puanları müdürlerin görevdeki başarısını öngörmek için kullandınız. Yeni rollerindeki performanslarının değerlendirmeleri artık erişilebilir durumda. Bu değerlendirmeler, sizin potansiyelleriyle ilgili klinik yargılarınıza ne kadar yakındır?

Bu örnek gerçek bir performans öngörüsü çalışmasına dayanıyor. O çalışmada katılımcı olsaydınız, muhtemelen sonuçlarından hoşnut kalmazdınız. Bu tür öngörülerde bulunmak için uluslararası bir danışmanlık şirketi tarafından işe alınan doktora seviyesinde psikologlar, performans değerlendirmeleriyle 0,15'lik bir korelasyona ulaştılar (YU = %55). Başka bir deyişle bir adayı diğerinden daha güçlü olarak puanlarken, –sizin Monica ve Nathalie'yle yaptığınız gibi– tercih ettikleri adayın ileride daha yüksek bir performans puanına ulaşması olasılığı %55, yani şans eserinden bir nebze daha iyiydi. En basit ifadeyle bu, hiç de etkileyici bir sonuç değil.

Belki de doğruluğun size gösterilen puanların öngörü açısından işe yaramaz olması nedeniyle düşük olduğunu düşünebilirsiniz. Bu yüzden adayların puanlamalarının aslında ne kadar işe yarar öngörücü bilgi içerdiğini sormalıyız. Bu puanlamalar performansla olası en yüksek korelasyona ulaşacak öngörücü bir skorda nasıl birleştirilebilir?

Bu soruların cevabı standart bir istatistiki yöntemde bulunabilir. Mevcut çalışmada bu yöntem 0,32'lik bir ideal korelasyon (YU = %60) veriyor; etkileyici olmaya uzak ama klinik öngörülerin eriştiğinden hatırı sayılır miktarda yüksek.

Çoklu regresyon adı verilen bu teknik, kestiricilerin ağırlıklı ortalaması olan kestirimci bir puan üretir. Bileşik öngörü ve hedef değişken arasındaki korelasyonu en üst seviyeye çıkarmak için seçilmiş ideal ağırlık setlerini bulur. Bu ideal ağırlıklar, öngörülerin OKH'sini (ortalama karesel hatasını) en alt seviyeye indirir; bu, istatistikte en küçük kareler prensibinin baskın rolüne çok uygun ve önemli bir örnektir. Tahmin edebileceğiniz gibi hedef değişkenle en yakın bağıntılı kestirici büyük bir ağırlık alır, işe yaramayan kestiricilerin ağırlığı ise sıfırdır. Ağırlıklar eksi de olabilir: Örneğin adayın ödenmemiş trafik cezalarının sayısı yöneticilik başarısı kestiricisi olarak muhtemelen eksi bir ağırlık alır.

Çoklu regresyon kullanımı bir *mekanik öngörü* örneğidir. Basit kurallardan ("işe liseyi bitirmiş herhangi birini al") sofistike yapay zekâ modellerine kadar çeşitlilik gösteren birçok mekanik öngörü türü vardır. Ancak en yaygın olanı doğrusal regresyon modelleridir ("yargı ve karar alma araştırmalarının yük beygiri" olarak anılırlar). Jargonu en alt seviyeye çekmek için doğrusal modellerden *basit modeller* olarak bahsedeceğiz.

Monica ve Nathalie'nin resmedildiği çalışma, klinik ve mekanik öngörülerin basit bir yapıyı paylaşan çok sayıdaki karşılaştırmasından biriydi.

- Bir *hedef sonucu* (aynı insanların işteki değerlendirmelerini) öngörmek için bir grup *kestirici değişkeni* (örneğimizde adayların puanlamaları) kullanılır.
- İnsan karar vericiler *klinik öngörü*lerde bulunurlar.
- Bir kural (çoklu regresyon gibi) aynı kestiricileri kullanarak aynı sonuçların *mekanik öngörüler*ini üretir.
- Klinik ve mekanik öngörülerin toplam isabetliliği karşılaştırılır.

Meehl: İdeal Model Sizi Yener

İnsanlar klinik ve mekanik öngörülerle karşılaştıklarında, birbirleriyle nasıl kıyaslandıklarını bilmek isterler. İnsan yargısı bir formüle kıyasla ne kadar iyidir?

Bu soru daha önce de sorulmakla birlikte asıl dikkati, Minnesota Üniversitesi'nde psikoloji profesörü olan Paul Meehl'in 1954 yılında *Clinical Versus Statistical Prediction: A Theoritical Analysis and a Review of the Evidence* adlı kitabını yayımlamasıyla çekti. Meehl akademik başarı ve psikiyatrik prognoz gibi sonuçlar için klinik bir yargının mekanik bir öngörüyle çarpıştırıldığı yirmi çalışmayı gözden geçirdi. Güçlü bir şekilde, basit mekanik kuralların genel olarak insan yargısından üstün olduğu kanısına ulaştı. Meehl klinik tedavi uzmanlarının ve diğer profesyonellerin sıklıkla kendilerine özgü bir güç olarak gördükleri şeyde, bilgiyi bütünleme becerilerinde, can sıkıcı bir şekilde zayıf olduklarını ortaya çıkardı.

Bu bulgunun ne kadar şaşırtıcı olduğunu ve gürültüyle ilgisini kavrayabilmek için, basit bir mekanik öngörü modelinin nasıl çalıştığını anlamalısınız. Mekanik öngörü modelinin tanımlayıcı özelliği, aynı kuralın bütün vakalara uygulanmasıdır. Her kestiricinin bir ağırlığı vardır ve o ağırlık vakadan vakaya değişmez. Bu katı kısıtlamanın modellerde insan yargılarına kıyasla büyük dezavantaj

oluşturduğunu düşünebilirsiniz. Bizim örneğimizde Monica'nın motivasyon ve teknik beceriler kombinasyonunun önemli bir artı olacağını ve diğer alanlardaki kısıtlamalarını dengeleyeceğini düşünmüş olabilirsiniz. Ve belki de diğer güçlü yanları göz önüne alındığında, Nathalie'nin bu iki alandaki zayıflıklarının ciddi bir sorun olmayacağını da düşündünüz. Üstü örtülü bir şekilde, iki kadın için farklı başarı yolları hayal ettiniz. Bu makul klinik spekülasyonlar iki vakada aynı kestiricilere fiilî olarak farklı ağırlıklar verir: basit bir modelin erişemeyeceği incelikte bir ayrıntı.

Basit modelin bir diğer sıkıntısı, bir kestiricide 1 birimlik bir artışın daima aynı etkiyi (ve 2 birimlik bir artışın etkisinin yarısını) yaratmasıdır. Klinik sezgi bu kuralı sıklıkla çiğner. Örneğin Nathalie'nin iletişim becerilerinden aldığı 10 tam puandan etkilendiyseniz ve bu puanın kestiriminizde bir yükselmeyi hak ettiğine karar verdiyseniz, basit bir modelin yapmayacağı bir şeyi yapmışsınız demektir. Bir ağırlıklı ortalama formülünde 10 puan ve 9 puan arasındaki fark, 7 ve 6 arasındaki farkla aynı olmalıdır. Klinik yargı bu kurala uymaz. Bunun yerine aynı farkın bir bağlamda sonuç üstünde önemsiz, diğerinde ise kritik olabileceği sezgisini yansıtır. Kontrol etmek isteyebilirsiniz ama hiçbir basit modelin Monica ve Nathalie'yle ilgili yargınızı tam anlamıyla açıklayabileceğini sanmıyoruz.

Bu vakalar için kullandığımız çalışma, Meehl'in örüntüsünün açık bir örneğiydi. Belirttiğimiz gibi klinik öngörüler, iş performansında 0,15'lik bir korelasyon (YU = %55) elde ettiler ama mekanik öngörü 0,32'lik bir korelasyona ulaştı (YU = %60). Monica ve Nathalie örneklerinin göreceli artılarında yaşadığınız güveni hatırlayın. Meehl'in sonuçları yargınızın kalitesi konusunda hissettiğiniz tatminin bir yanılsama yani *geçerlilik yanılsaması* olduğunu şiddetle savunuyor.

Geçerlilik yanılsaması, yaygın olarak öngörü görevinin iki aşamasının (vakaları erişilebilir kanıtlara göre değerlendirmek ve

gerçek sonuçları kestirmek) birbirinden ayırt edilememesi nedeniyle, öngörücü yargıların verildiği her yerde bulunur. Adaylardan hangisinin daha iyi *göründüğü* konusunda değerlendirmenizden emin olabilirsiniz ama gerçekte hangisinin daha iyi *olacağını* tahmin etmek bambaşka bir karmaşık meseledir. Örneğin Nathalie'nin Monica'dan daha güçlü bir aday gibi göründüğünü öne sürmek güvenlidir ama Nathalie'nin Monica'dan daha başarılı bir üst düzey yönetici olacağını öne sürmek hiç de güvenli değildir. Nedeni çok açık: İki vakayı değerlendirmek için bilmeniz gerekenin büyük kısmını biliyorsunuz ama geleceğe bakmak fazlasıyla belirsizdir.

Ne yazık ki bu fark, düşüncemizde bulanıklaşır. Vakalar ve öngörüler arasındaki ayrımın kafanızın karıştırdığını düşünüyorsanız, yalnız değilsiniz. Bu ayrımı herkes kafa karıştırıcı bulur. Ancak öngörülerinizde vaka değerlendirmeleriniz kadar eminseniz, geçerlilik yanılsamasının kurbanısınızdır.

Klinik tedavi uzmanları da geçerlilik yanılsamasından muaf değildirler. Klinik psikologların Meehl'in, sıradan formüllerin, sürekli olarak uygulandıklarında, klinik yargıya üstün geldiği bulgusuna verdiği tepkiyi hayal edebilirsiniz. Tepki, şok, inanmama ve klinik sezginin olağanüstü yanlarını inceler gibi davranan sığ araştırmaya karşı küçümsemeyi birleştirdi. Bu kolay anlaşılır bir tepki. Meehl'in örüntüsü öznel yargı deneyimiyle çelişiyor ve çoğumuz kendi deneyimimize bir âlimin iddiasından daha fazla güveniriz.

Meehl'in kendisi de bulguları konusunda karışık duygular içindeydi. Adı istatistiğin klinik yargıya üstünlüğüyle bağdaştırıldığı için, onu insafsız bir insan içgörüsü eleştirmeni ya da nicel olan her şeyin koruyucu babası olarak hayal edebiliriz. Ama bu bir karikatür olurdu. Meehl akademik kariyerine ek olarak faal bir psikanalistti. Ofisinde Freud'un bir fotoğrafı asılıydı. Sınıflara psikolojinin yanı sıra, felsefe ve hukuk konusunda dersler veren, metafizik, din, politika bilimi ve hatta parapsikoloji ("telepatide gerçeklik payı olduğunda" ısrarcıydı) hakkında yazan, birçok alanda

bilgili bir âlimdi. Bu özelliklerin hiçbiri dediğim dedik bir sayı adamı klişesine uymuyor. Meehl'in klinik tedavi uzmanlarıyla bir husumeti yoktu. Ancak kendi ifadesiyle, girdileri birleştirmeye mekanik yaklaşımın lehine olan kanıtlar "çok büyük ve sürekliydi".

"Büyük ve sürekli" adil bir tanım. 2000 yılında 136 çalışmanın gözden geçirilmesi, mekanik birleştirmenin klinik yargıya üstün geldiğini tartışmasız bir şekilde doğruladı. Makalede incelenen araştırma sarılık teşhisi, askerî hizmete uygunluk ve evlilikten tatminin de aralarında olduğu geniş bir konu yelpazesini kapsıyordu. Çalışmaların 63'ünde mekanik öngörü daha doğruydu; 65'inde istatistiki beraberlik sağlandı ve 8 vakada yarışı klinik öngörü kazandı. Bu sonuçlar aynı zamanda klinik yargıdan daha hızlı ve ucuz olan mekanik öngörünün avantajlarını olduğundan daha hafif gösteriyor. Dahası bu çalışmaların pek çoğunda insan yargıçların haksız bir avantajı vardı çünkü bilgisayar modeline sağlanmayan "özel" bilgilere erişebilmişlerdi. Bulgular açık bir sonucu destekliyor: *Basit modeller insanları yener.*

Goldberg: Kendi Modeliniz Sizi Yener

Meehl'in keşfi önemli sorular doğuruyor. Formül tam olarak neden üstün? Formül neyi daha iyi yapıyor? Hatta insanların neyi daha kötü yaptığı sorusu daha iyi bir soru olabilir. Cevap insanların pek çok açıdan istatistiki modellerden daha aşağıda olması. Kritik zayıflıklarından biri, gürültülü olmaları.

Bu çıkarımı desteklemek için, basit modeller üstünde yapılmış ve Oregon'da küçük bir şehir olan Eugene'de başlayan farklı bir araştırma akımına dönüyoruz. Paul Hoffman akademiye karşı sabırsız olan varlıklı ve vizyon sahibi bir psikologdu. Olağanüstü verimli birkaç araştırmacıyı aynı çatı altına topladığı bir araştırma enstitüsü kurdu. Enstitü Eugene'i insan yargısının incelenmesi konusunda dünyaca ünlü bir merkeze dönüştürdü.

Söz konusu araştırmacılardan biri, Büyük Beş Kişilik Özelliği modelinin geliştirilmesindeki öncü rolüyle tanınan Lewis Goldberg idi. Goldberg 1960'ların sonunda, Hoffman'ın daha önce yaptığı çalışmaların izinden giderek, bireyin yargılarını tanımlayan istatistik modellerini inceledi.

Böyle bir yargıç modeli inşa etmek, bir gerçeklik modeli inşa etmek kadar kolaydır. Aynı kestiriciler kullanılır. İlk örneğimizde kestiriciler bir müdürün performansının beş puanlaması idi. Aynı araç, yani çoklu regresyon kullanılır. Tek fark hedef değişkendir. Bir gerçek sonuç seti öngörmek yerine, bir yargı grubunun (örneğin Monica, Nathalie ve diğer müdürler hakkında *sizin* yargılarınızın) öngörülmesi için formül uygulanır.

Yargılarınızı ağırlıklı bir ortalama olarak model almak büsbütün tuhaf gelebilir çünkü görüşlerinizi bu şekilde oluşturmazsınız. Monica ve Nathalie'yi klinik olarak düşündüğünüzde, iki vakaya aynı kuralı uygulamadınız. Hatta hiç kural uygulamadınız. Yargıç modeli, bir karar alıcının gerçek hayatta yargıya varma şeklinin gerçekçi bir tarifi değildir.

Ancak gerçekten doğrusal bir formül hesaplamasanız bile, kararlarınızı yine de formülle hesaplamış gibi verebilirsiniz. Usta bilardocular, hiç böyle bir şey yapmasalar da belli bir atışın mekaniğini tarif eden karmaşık denklemleri çözmüş gibi davranırlar. Benzer bir şekilde, aslında yaptığınız şey çok daha karmaşık olsa da basit bir formül kullanmışsınız gibi kestirici üretebilirsiniz. İnsanların makul bir doğrulukla ne yapacaklarını öngören bir -miş gibi modeli, sürecin bir tanımı olarak bariz bir şekilde hatalıyken bile, işe yarar. Basit karar modelleri için de durum budur. Karar çalışmalarının kapsamlı bir incelemesi 237. çalışmada yargıç modeli ve yargıcın klinik kararları arasındaki ortalama korelasyonun 0,80 (YU = 79) olduğunu ortaya çıkardı. Kusursuz olmaya çok uzak olmakla birlikte, bu korelasyon bir -miş gibi teorisini desteklemeye yetecek kadar yüksektir.

Goldberg'in araştırmasına itici güç olan soru, yargıca ilişkin basit bir modelin, gerçek sonuçları ne kadar iyi öngöreceğiydi. Model yargıcın kaba bir yaklaşığı olduğu için, akla yatkın bir şekilde, eşit derecede iyi performans sergilemeyeceğini varsayabiliriz. Yargıcın yerini model aldığı zaman, isabetliliğin ne kadarı kaybedilir?

Cevap sizi şaşırtabilir. Öngörüler, model tarafından üretildiğinde doğruluklarından kayıp verilmedi. İyileştiler. Çoğu örnekte model, dayandığı profesyonelden daha iyi öngörülerde bulundu. Taklidi orijinal üründen daha iyiydi.

Bu çıkarım pek çok alanda birçok çalışmayla onaylandı. Goldberg'in çalışmasının eski bir taklidi, lisansüstü eğitim başarısıyla ilgili öngörüleri içeriyordu. Araştırmacılar doksan sekiz katılımcıdan on ipucuyla doksan öğrencinin not ortalamasını tahmin etmelerini istediler. Araştırmacılar, bu tahminlere dayanarak her katılımcının yargılarının doğrusal modelini oluşturdular ve katılımcıların ve katılımcıların modellerinin not ortalamasını ne kadar doğru öngördüklerini karşılaştırdılar. Doksan sekiz katılımcının her biri için model, katılımcıdan daha iyi iş çıkardı! Onlarca yıl sonra, elli yıllık araştırmanın gözden geçirilmesi, yargıçların modellerinin, sürekli ve tutarlı şekilde, modelledikleri yargıçtan daha iyi performans sergilediği sonucunu ortaya koydu.

Bu çalışmalardaki katılımcıların performansları hakkında kişisel geri besleme alıp almadıklarını bilmiyoruz. Ama biri size yargılarınızın kaba bir modelinin –neredeyse bir karikatürünün– sizden daha isabetli olduğunu söylese hoşunuza gitmeyeceğini hayal edebiliyorsunuzdur. Çoğumuz için karar aktivitesi, tam olarak basit kurallara uymaması nedeniyle karmaşık, zengin ve ilginçtir. Kendimiz ve karar alma becerimiz hakkında en iyi hissettiğimiz zaman, karmaşık kurallar icat ettiğimiz ve onları uyguladığımız ya da bir vakayı diğerlerinden farklı kılan bir içgörüye sahip olduğumuz, kısacası sade bir ağırlıklı ortalama işlemine indirgenemeyecek kararlar verdiğimiz zamandır. Yargıcın modeli

çalışmaları Meehl'in inceliğin büyük ölçüde ziyan olduğu sonucunu pekiştirmiştir. Karmaşıklık ve zenginlik genel olarak daha doğru öngörülere yol açmaz.

Neden böyledir? Goldberg'in bulgularını anlamak için, önce sizinle modeliniz arasındaki farkların açıklamasını anlamalıyız. Gerçek kararlarınızla ve onları öngören basit bir modelin çıktısı arasındaki uyuşmazlıklara ne sebep olur?

Kararlarınızın istatistiksel bir modeli, içerdikleri bilgiye herhangi bir şey ekleyemez. Modelin bütün yapabildiği çıkarmak ve basitleştirmektir. Özellikle, yargılarınızın basit modeli sizin sürekli ve tutarlı bir şekilde uyduğunuz karmaşık kuralların hiçbirini temsil etmez. İletişim becerisi puanlamasında 10 ve 9 arasındaki farkın, 7 ve 6 arasındaki farktan daha önemli olduğunu ya da bütün boyutlarda sağlam bir 7 alan çok yönlü bir adayın aynı ortalamaya belirgin güçlü ve zayıf yanlarla ulaşan bir diğer adaya tercih edilmesi gerektiğini düşünüyorsanız, –modeliniz siz onları kusursuz bir süreklilikle uyguluyor olsanız bile– karmaşık kurallarınızı yeniden üretmeyecektir.

İncelikli kurallarınızın yeniden üretilememesi, inceliğinizin geçerli olduğu durumda bir doğruluk kaybına neden olacaktır. Örneğin iki girdi (beceri ve motivasyon) üzerinden zor bir görevin başarılı olup olmayacağını tahmin etmeniz gerektiğini varsayalım. Ağırlıklı ortalama iyi bir formül olmaz çünkü motivasyon ne kadar yüksek olursa olsun, ciddi bir beceri eksikliğini aşmaya yetmeyecektir. Ya da tam tersi. İki girdinin daha karmaşık bir kombinasyonunu kullanırsanız, kestirimci doğruluğunuz artar ve bu inceliği yakalamayı başaramayacak bir modelin erişebileceğinden daha yüksek olur. Öte yandan karmaşık kurallar genellikle size geçerlilik yanılsaması verir ve aslında kararlarınızın kalitesine zarar verir. Bazı incelikler geçerlidir ama hepsi değil.

Ek olarak, basit bir modeliniz kararlarınızdaki örüntü gürültüsünü temsil etmeyecektir. Belli bir vakaya verdiğiniz keyfî

tepkilerden doğan olumlu ve olumsuz hataları taklit edemez. Model ayrıca anlık bağlamın ve o kararı verdiğiniz zamanki ruhsal hâlinizin etkilerini de yansıtamaz. Büyük olasılıkla, bu gürültülü yargı hataları herhangi bir şeyle sistematik olarak bağıntılı değildir, bu, pek çok amaç açısından rastgele sayılabilecekleri anlamına gelir.

Kararlarınızdan gürültüyü çıkarma etkisi her zaman kestirimci isabetliliğinizde iyileşmeye neden olacaktır. Örneğin, öngörülerinizle sonuç arasındaki korelasyonun 0,50 (YU = %67) olduğunu ama kararlarınızın değişkenliğinin % 50'sinin gürültüden oluştuğunu varsayalım. Kararlarınız –bir modelde olabileceği gibi– gürültüden arındırılabilseydi aynı sonuçla korelasyonları 0,71'e (YU = %75) sıçrardı. Gürültüyü mekanik olarak azaltmak öngörücü yargının geçerliliğini artırır.

Kısaca, yerinize bir modelinizi koymak iki şey yapar: İnceliğinizi ve örüntü gürültünüzü bertaraf eder. Yargıcın modelinin yargıçtan daha geçerli olduğu bulgusu önemli bir mesaj içerir: İnsan kararlarında incelikli kurallardan sağlanan kazanç –eğer varsa– genellikle gürültünün yıkıcı etkileri telafi etmeye yetmez. Daha incelikli, daha içgörülü ve düşüncenizin doğrusal karikatüründen daha farklı olduğunuza inanabilirsiniz. Ama aslında çoğunlukla daha gürültülüsünüzdür.

Karmaşık öngörü kuralları, geçerli içgörülerden faydalandıkları duygusunu taşımamıza rağmen, neden doğruluğa zarar verirler? Öncelikle, insanların icat ettiği karmaşık kuralların çoğu genel olarak doğru olmama eğilimindedir. Ama bir sorun daha var: Karmaşık kurallar prensipte geçerli olduğunda bile, kaçınılmaz olarak nadiren gözlenen şartlarda geçerli olurlar. Örneğin, diğer boyutlardaki puanları vasat bile olsa, olağanüstü orijinal adayların işe alınmayı hak ettiğine karar verdiniz. Sorun, olağanüstü orijinal adayların, tanım gereği, olağanüstü nadir bulunmasıdır. Orijinalliğin değerlendirilmesinin güvenilmez olması yüksek ihtimal olduğu için, bu ölçüdeki pek çok yüksek puan şans eseri

yakalanır ve gerçek anlamda orijinal yetenek sıklıkla saptanamadığıyla kalır. "Orijinallerin" süper star olabileceğini onaylayabilecek performans değerlendirmeleri de kusurludur. İki uçtaki ölçüm hataları kaçınılmaz olarak öngörülerin geçerliliğini zayıflatır ve nadir olayların gözden kaçması özellikle olasıdır. Gerçek inceliğin avantajları ölçüm hatasında hızla boğulur.

Martin Yu ve Nathan Kuncel'in bir çalışması Goldberg'in sunumunun daha radikal bir versiyonunu bildirdi. Bu çalışma (Monica ve Nathalie örneğinin temelini oluşturan çalışma) üç ayrı örnekte 847 adayı üst düzey pozisyonlar için değerlendirmek için uzmanlar istihdam eden uluslararası bir danışmanlık şirketinin verilerini kullandı. Uzmanlar sonuçları yedi farklı değerlendirme boyutunda puanladılar ve her birine toplam bir kestirimci puan vermek için klinik yargılarını kullandılar. Ortaya pek etkileyici olmayan sonuçlar çıktı.

Yu ve Kuncel yargıçları kendilerinin en iyi basit modeliyle değil, *rastgele* bir doğrusal modelle karşılaştırmaya karar verdiler. Yedi kestirici için on bin rastgele ağırlık seti ürettiler ve iş performansını öngörmek için on bin rastgele formül uyguladılar.

Çarpıcı bulguları, herhangi bir doğrusal modelin vakaların tamamına istikrarlı bir şekilde uygulandığında, aynı bilgiden bir sonuç öngörmede insan karar alıcılardan daha iyi iş çıkarmasının muhtemel olduğuydu. Üç örnekten birinde, on bin rastgele ağırlıklandırılmış doğrusal modelin %77'si insan uzmanlardan daha iyi sonuç aldı. Diğer iki örnekte, rastgele modellerin %100'ü insanlardan daha iyi performans sergiledi. Ya da en açık ifadesiyle, o çalışmada uzmanlardan daha kötü bir iş çıkaracak basit bir model üretmek neredeyse imkânsızdı.

Bu araştırmanın sonucu Goldberg'in yargıcın modeli üstündeki çalışmasından aldığımız sonuçtan daha güçlüdür, ki bu uç bir örnektir. Bu ortamda, insan yargıçlar mutlak şekilde kötü performans sergilediler; bu durum, etkileyicilikten uzak doğrusal modellerin

bile onları geçmesini açıklamaya yardımcı olacaktır. Elbette her modelin her insanı alt edeceği sonucunu çıkarmamalıyız. Yine de basit bir kurala mekanik olarak uymanın (Yu ve Kuncel buna "düşüncesiz tutarlılık" diyor) zor bir problemde yargıyı önemli ölçüde iyileştirebilmesi gürültünün klinik öngörülerin geçerliliği üstündeki büyük etkisini ortaya koymaktadır.

Bu hızlı tur gürültünün klinik yargıyı nasıl sakatladığını gösteriyor. Öngörücü yargılarda, basit formüller –gerçekliğin modelleri, bir karar alıcının modelleri, hatta rastgele üretilmiş modeller– insan uzmanları kolayca alt edebiliyor. Bu bulgu bir sonraki bölümün konusu olan gürültüden arındırılmış yöntemlerin kullanımını destekliyor.

Yargılardan ve Modellerden Bahsederken

"İnsanlar yargıda bulunurken karmaşıklığı kavradıklarına ve incelik kattıklarına inanır. Ancak karmaşıklık ve incelik çoğunlukla ziyan olur; genellikle basit modellerin doğruluğuna bir şey eklemezler."

"Paul Meehl'in kitabının yayımlanmasından altmış sene sonra mekanik öngörünün insanlardan üstün olması hâlâ şoke etmeye devam ediyor."

"Yargıda bulunmada o kadar çok gürültü var ki yargıya varan birinin gürültüden arındırılmış modeli fiilen yargıya varandan daha doğru kestirimlere ulaşabiliyor."

BÖLÜM 10

Gürültüsüz Kurallar

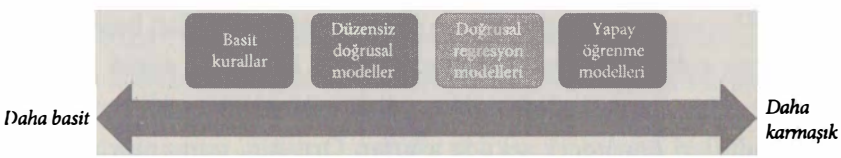
Son yıllarda yapay zekâ (YZ), özellikle yapay öğrenme teknikleri makinelerin daha önce tamamen insana özgü olarak görülen pek çok görevi gerçekleştirmesine olanak sağladı. Yapay öğrenme algoritmaları yüzleri tanıyabiliyor, dilleri tercüme edebiliyor ve radyoloji imgelerini okuyabiliyor. Aynı anda binlerce sürücüye şaşırtıcı bir hız ve doğrulukla yol tarifi üretme gibi hesaplama sorunlarını çözebiliyorlar. Ayrıca zor öngörü görevlerini de gerçekleştiriyorlar: Yapay öğrenme algoritmaları ABD Yüksek Mahkemesi'nin kararlarını öngörebiliyor, hangi sanıkların kefaletle salındıktan sonra ortadan kaybolacağını saptıyor ve çocuk koruma hizmetlerine gelen çağrıların hangilerinin daha acil bir sosyal hizmet görevlisi ziyaretini gerektirdiğini değerlendiriyorlar.

Her ne kadar bugünlerde *algoritma* sözcüğünü duyduğumuzda aklımıza gelen uygulamalar bunlar olsa da terimin daha geniş bir anlamı var. Bir sözlükte algoritma, "özellikle bilgisayar tarafından yapılan hesaplamalarda ya da diğer sorun çözme işlemlerinde uyulması gereken bir süreç ya da kural grubu" olarak geçiyor. Bu tanıma göre bir önceki bölümde değindiğimiz basit modeller ve diğer mekanik karar biçimleri de algoritmadır.

Aslında neredeyse gülünecek kadar basit kurallardan en sofistike ve anlaşılması en güç makine algoritmalarına kadar pek çok mekanik yaklaşım türü, insan yargısından daha üstün performans sergileyebiliyor. Ve bu üstün performansın kilit bir nedeni –tek nedeni olmamakla birlikte– bütün mekanik yaklaşımların gürültüsüz olmasıdır.

Kurala dayalı yaklaşımların farklı türlerini incelemek ve her yaklaşımın nasıl ve hangi şartlar altında değerli olabileceğini öğrenmek için, yolculuğumuza Bölüm 9'daki modellerle –çoklu regresyona dayalı basit modellerle yani doğrusal regresyon modelleriyle– başlayacağız.

Bu başlangıç noktasından hareketle, karmaşıklık yelpazesinde iki aksi yönde hareket edeceğiz. Önce uç noktada basitliği arayacak, ardından üstüne daha fazla karmaşıklık ekleyeceğiz (Tablo 11).



ŞEKİL 11: *Dört tip kural ve algoritma*

Daha Fazla Basitlik: Sağlam ve Güzel

Robyn Dawes da Eugene, Oregon'un, 1960'lı ve 1970'li yıllarda yargı konusunu inceleyen yıldızlar takımın bir üyesiydi. 1974 senesinde Dawes, öngörü görevlerinin basitleştirilmesinde önemli bir başarı yakaladı. Düşüncesi şaşırtıcı, neredeyse bütün inanışlara tersti: Her kestiricinin net ağırlığını saptamak için çoklu regresyonu kullanmak yerine, bütün kestiricilere eşit ağırlık verecekti.

Dawes eşit ağırlık formülünü *düzensiz doğrusal model* olarak adlandırdı. Şaşırtıcı keşfi, bu eşit ağırlıklı modellerin de "düzenli" regresyon modelleri kadar doğru ve klinik yargılardan kat kat üstün olduğuydu.

Düzensiz modellerin destekçileri bile bu iddianın mantık dışı ve "istatistiki sezgiye ters" olduğunu kabul ediyorlar. Hatta Dawes ve asistanı Bernard Corrigan başlangıçta çalışmalarını bilimsel dergilerde yayımlanmakta çok zorlandılar; editörler onlara inanmıyordu. Bir önceki bölümdeki Monica ve Nathalie örneğini düşünürseniz, muhtemelen bazı kestiricilerin diğerlerinden daha önemli olduğuna inanıyorsunuzdur. Örneğin, çoğu insan liderliğe teknik becerilerden daha fazla ağırlık verirdi. Düz ağırlıksız bir ortalama, bir insanın performansını dikkatli bir şekilde ağırlıklı ortalamadan ya da bir uzmanın yargısından nasıl daha iyi öngörebilir?

Bugün Dawes'ın çığır açan bu keşfinden yıllar sonra, çağdaşlarını çok şaşırtan bu istatistiki fenomen artık çok iyi anlaşıldı. Bu kitabın daha önceki bölümlerinde açıklandığı gibi çoklu regresyon kare hataları, en düşük seviyeye çeken "ideal" ağırlıkları hesaplar. Ancak çoklu regresyon, *orijinal verideki* hatayı en düşük seviyeye çeker. Bu nedenle formül kendini verideki bütün bilinmeyen tesadüfleri öngörecek şekilde ayarlar. Örneğin, numune yüksek teknik becerilere sahip ve alakasız nedenlerden olağanüstü iyi performans sergilemiş birkaç yöneticiyi içeriyorsa model teknik becerinin ağırlığını abartacaktır.

Buradaki zorluk şudur: Formül *numune dışı* uygulandığında, yani farklı bir veri setinde sonuçları tahmin etmek için kullanıldığında, ağırlıklar ideal olmaktan çıkar. Orijinal numunedeki tesadüfler, özellikle tesadüf oldukları için artık yoktur; yeni numunede yüksek teknik becerilere sahip müdürlerin hepsi süper star değildir. Ve yeni numunenin formülün öngöremeyeceği farklı tesadüfleri vardır. Bir modelin kestirimci doğruluğunun doğru ölçümü yeni bir numunedeki performansıdır ve buna *çapraz doğrulanmış korelasyon* denir. Aslında bir regresyon modeli orijinal numunede *fazla* başarılıdır ve çapraz doğrulanmış korelasyon neredeyse her zaman orijinal veride olduğundan daha düşüktür. Dawes ve Cor-

rigan eşit ağırlık modellerini birçok durumda çoklu regresyon modelleriyle karşılaştırdılar (çapraz doğruladılar). Örneklerinden biri Illinois Üniversitesi'nde psikoloji eğitimi alan doksan lisansüstü öğrencisinin birinci yıl not ortalamalarının akademik başarıyla ilgili on değişken kullanılarak yapılmış tahminlerini içeriyordu: yatkınlık test puanları, lisans notları, çeşitli akran puanlamaları (dışa dönüklük gibi) ve çeşitli öz değerlendirmeler (vicdanlılık gibi). Standart çoklu regresyon modeli 0,69'luk bir korelasyon elde etti ve bu değer çapraz doğrulamada 0,57'ye (YU = %69) düştü. Birinci sınıf not ortalamasıyla eşit ağırlık modelinin korelasyonu da hemen hemen aynıydı: 0,60 (YU = %70). Daha pek çok çalışmada benzer sonuçlar elde edildi.

Çapraz doğrulamada doğruluk kaybı orijinal numune küçük olduğunda en kötü hâlini alır çünkü tesadüfler küçük numunelerde daha büyük görünürler. Dawes'ın dikkat çektiği sorun, sosyal bilim araştırmalarında kullanılan numunelerin genellikle sözde ideal ağırlıklandırma avantajı kaybolacak kadar küçük olmasıydı. İstatistikçi Howard Wainer'ın uygun ağırlıkların tahmini üzerine kaleme aldığı akademik makalenin alt başlığında, akılda kalacak şekilde belirttiği gibi: "Benim için Fark Etmez." ya da Dawes'ın kelimeleriyle, "ölçümlerimizden daha kesin modellere ihtiyacımız yok." Eşit ağırlık modelleri iyi iş çıkarırlar çünkü örnekleme kazalarına açık değildirler.

Dawes'ın çalışmasının öncelikli çıkarımı yaygın olarak bilinmeyi hak ediyor: Tahmin etmeye çalıştığınız sonuç hakkında ön verilere sahip olmadan da geçerli istatiksel öngörülerde bulunabilirsiniz. Tek ihtiyacınız sonuçla bağıntılı olduğuna güvenebileceğiniz bir kestirimciler koleksiyonudur.

Bölüm 9'daki örnekteki gibi bir dizi boyutta puanlanmış üst düzey yöneticilerin performansı hakkında öngörülerde bulunduğunuzu varsayın. Bu puanların önemli özellikleri ölçtüğüne güveniyorsunuz ama her puanın performansı ne kadar iyi öngördüğü konusunda

hiçbir veriniz yok. Çok sayıda müdürden oluşan bir numunenin performansını izlemek için birkaç yıl bekleme lüksünüz de yok. Buna rağmen yedi puanı alıp onları eşit olarak ağırlıklandırmak için gereken istatistik çalışmasını yapabilir ve sonucu öngörünüz olarak kullanabilirsiniz. Bu eşit ağırlık modeli ne kadar iyi olurdu? Sonuçla korelasyonu 0,25 (YU = %58) olurdu. Yani korelasyonu 0,15 (YU = %55) olan klinik öngörülerden çok daha üstün ve elbette bir çapraz doğrulanmış regresyon modeline hayli yakın.

Dawes'ın yargı öğrencileri arasında yaygın bir düşünce hâline gelen ifadesiyle, eşit ağırlıkta "sağlıklı bir güzellik" var. Bu fikri ortaya süren ufuk açıcı makalenin son cümlesi bir diğer güçlü ve anlamlı özeti sunuyor: "Bütün hile hangi değişkenlere bakacağınıza karar vermek ve nasıl toplama yapacağınızı bilmekte saklı."

Daha da Basitlik: Basit Kurallar

Bir diğer basitleştirme tarzı *tutumlu modeller* ya da *basit kurallardır.* Tutumlu modeller saçmalık derecesinde basitleştirilmiş, zarfın arka yüzünde yapılmış hesaplamalar gibi görünen gerçeklik modelleridir. Ancak bazı ortamlarda şaşırtıcı derecede iyi tahminler üretebilirler.

Bu modeller çoklu regresyonun çoğu insanın şaşırtıcı bulduğu bir özelliği üstüne inşa edilir. Diyelim sonucu kestirmede güçlü iki kestirici kullanıyorsunuz: Sonuçla korelasyonları 0,60 (YU = % 71) ve 0,55 (YU = %69). Ayrıca iki kestiricinin birbirleriyle de 0,50'lik bir korelasyonla bağıntılı olduğunu varsayın. İki kestirici en uygun şekilde birleştirilmişken, öngörünüzün ne kadar iyi olacağını tahmin edersiniz? Cevap hayal kırıklığına uğratacak türden: Korelasyon 0,67 (YU = %73). Daha öncesinden yüksek ama çok fazla değil.

Bu örnek genel bir kuralı temsil ediyor: İki ya da daha fazla bağıntılı kestiricinin kombinasyonu, en iyilerinin tek başına ol-

duğundan sadece biraz daha fazla kestirimcidir. Gerçek hayatta kestiriciler neredeyse her zaman birbiriyle bağıntılı olduğu için, bu istatistiksel olgu tahminde az sayıda kestiriciden faydalanan tutumlu yaklaşımların kullanılmasını destekler. Az ya da sıfır hesaplamayla uygulanabilen basit kurallar bazı ortamlarda, çok daha fazla kestirici kullanan modellere kıyasla etkileyici düzeyde doğru tahminler üretmişlerdir.

Bir grup araştırmacı 2020 yılında, çeşitli tahmin sorunlarına tutumlu bir yaklaşım uygulamak için büyük ölçekli bir çalışma yayımladılar. Söz konusu sorunlardan biri de kefalet yargıçlarının sanıkları duruşmaya kadar alıkoyma ya da salma tercihiydi. Bu karar sanığın davranışının üstü örtülü bir öngörüsüdür. Hatalı bir şekilde kefaletle salınmayan bir kişi hem bireye hem topluma hatırı sayılır bir maliyet yaratacak şekilde gereksiz yere alıkonulur. Kefalet hakkı yanlış bir sanığa tanındığı zaman ise kişi duruşma öncesi kaçabilir, hatta bir suç daha işleyebilir.

Araştırmacıların inşa ettiği model, bir sanığın kefaletle serbest bırakılması durumunda kaçma olasılığını öngörmek konusunda güçlü olarak bilinen iki girdi kullanıyor: sanığın yaşı (ileri yaştaki insanların kaçma riski daha düşüktür) ve geçmişte atlanmış mahkeme tarihlerinin sayısı (daha önce mahkeme tarihini atlayan insanlar bunu tekrarlama eğiliminde olur). Model bu iki girdiyi risk skoru olarak kullanılabilecek bir grup puana dönüştürüyor. Bir sanık için risk hesabı bırakın bilgisayarı hesap makinesi bile gerektirmiyor.

Gerçek bir veri grubuna karşı test edildiğinde, bu tutumlu model çok daha fazla sayıda değişken kullanan istatistik modelleri kadar iyi sonuç verdi. Tutumlu model bütün kefalet yargıçlarının sanığın kaçma riskini öngörmede olduklarından daha başarılı oldu.

Aynı tutumlu yaklaşım, küçük tam sayılarla (-3 ve +3 arası) ağırlıklandırılmış beş özelliğe kadar kullanan hâliyle, mamografi bilgisinden bir tümörün ciddiyetini saptama, bir kalp hastalığını

teşhis etme ve kredi riskini öngörme gibi çeşitli görevlere uygulandı. Bütün bu görevlerde, tutumlu kural daha karmaşık regresyon modelleri kadar iyi iş çıkardı (her ne kadar yapay öğrenme kadar iyi olmasa da).

Basit kuralların gücünü gösteren bir başka örnekte başka bir araştırma ekibi, benzer ama ayrı bir yargısal sorunu inceledi. Sadece iki girdi kullanarak, bir sanığın risk düzeyini değerlendirmek için 137 değişken kullanan mevcut bir aletin geçerliliğini karşılayabildiler. Şaşırtıcı olmayan bir şekilde, bu iki kestirici (yaş ve daha önce aldığı hüküm sayısı) kefalet modelinde kullanılan iki faktörle yakından ilişkiliydi ve suç davranışıyla bağlantıları daha önceden kanıtlarla sabitlenmişti.

Tutumlu kuralların cazibesi şeffaf ve kolay uygulanabilir olmalarıdır. Dahası bu avantajlar doğruluk bakımından, daha karmaşık modellere kıyasla nispeten daha az maliyetle elde edilir.

Daha Fazla Karmaşıklık: Yapay Öğrenmeye Doğru

Yolculuğumuzun ikinci kısmında, şimdi karmaşıklık yelpazesinin diğer yönüne doğru ilerleyelim. Ya daha fazla kestirici kullanabilsek, her biri hakkında daha fazla veri toplasak, hiçbir insanın saptayamayacağı ilişki örüntüleri yakalasak ve bu örüntüleri daha iyi öngörüye ulaşabilmek için modellesek? Özünde, yapay zekânın vaadi budur.

Karmaşık analizler için çok büyük veri setleri gereklidir ve bu tür veri setlerinin erişilebilirliğinin artması YZ'nin son yıllardaki hızlı ilerlemesinin ana sebeplerinden biridir. Örneğin, büyük veri setleri mekanik olarak *kırık bacak istisnalarıyla* baş etmeyi mümkün kılar. Bu neredeyse şifreli ifade Meehl'in hayal ettiği bir örneğe dayanır: Bu akşam insanların sinemaya gitme olasılığını öngörmek için tasarlanmış bir model düşünün. Modele duyduğunuz güvenden bağımsız olarak, şans eseri belli bir insanın

biraz önce bacağını kırdığını bilirseniz, o kişinin akşamının neye benzeyeceğini muhtemelen modelden daha iyi bilirsiniz.

Basit modeller kullanırken, kırık bacak prensibi karar alıcılar için önemli bir ders barındırır: onlara ne zaman modeli geçersiz kılmaları ne zaman kılmamaları gerektiğini söyler. Modelin değerlendirmeye alamayacağı kesin bilginiz varsa, ortada gerçek bir kırık bacak varsa, modelin önerisini geçersiz kılmalısınız demektir. Öte yandan, bazen bu tür özel bilgilere sahip olmadan da modelin önerisine karşı çıkarsınız. Bu durumlarda, modeli geçersiz kılma arzunuz aynı kestiricilere uyguladığınız kişisel bir örüntüyü yansıtır. Bu kişisel örüntünün geçersiz olması çok olası olduğu için, modelinizi geçersiz kılmaktan sakınmalısınız; müdahalenizin tahmini daha az doğru kılması çok mümkündür.

Öngörü görevlerinde yapay öğrenme modellerinin başarısının nedenlerinden biri de bu tür kırık bacakları –insanların düşündüğünden çok daha fazlasını– keşfetmeye muktedir olmalarıdır. Çok sayıda vakayla ilgili çok büyük miktarda veri sayesinde, sinema izleyicilerinin davranışını izleyen bir model, örneğin düzenli sinema günlerinde hastaneye giden insanların o akşam film izleme ihtimalinin azaldığını öğrenebilir. Nadir olayların tahminini bu şekilde iyileştirmek insan gözetimine ihtiyacı azaltır.

Yapay zekânın yaptığı şey sihir ve idrak içermez; sadece örüntü bulmaktan ibarettir. Yapay öğrenmenin gücüne hayranlık duymalıyız ama öte yandan bir yapay zekânın bacağı kırılan bir insanın sinema akşamını *neden* atlayacağını anlamasının muhtemelen zaman alacağını unutmamalıyız.

Bir Örnek: Daha İyi Kefalet Kararları

Daha önce bahsettiğimiz araştırma ekibinin kefalet kararlarına basit kurallar uyguladığı dönemde, Sendhil Mullainathan'ın liderliğindeki bir başka ekip sofistike YZ modellerini aynı görevi

yapmak için eğitti. YZ ekibi daha büyük bir veri grubuna, 758.027 kefalet kararına erişebildi. Ayrıca ekibin her vakada yargıca açık olan başka bilgilere de erişimi vardı: sanığın hâlihazırdaki suçu, sabıka kaydı ve daha önceki mahkemeye katılmama sicili. Algoritmayı eğitmek için yaş dışında hiçbir demografik veri kullanılmadı. Araştırmacılar ayrıca, her vaka için, sanığın salınıp salınmadığını ve salındıysa mahkemeye gelmemezlik edip etmediğini ve yeniden tutuklanıp tutuklanmadığını da biliyorlardı. (Sanıkların %74'ü salınmıştı ve bu %74'ün %15'i mahkemeye gelmemiş ve %26'sı yeniden tutuklanmıştı.) Araştırmacılar bir yapay öğrenme algoritmasını bu verilerle eğittiler ve performansını değerlendirdiler. Model bir yapay öğrenme üzerinden inşa edildiği için doğrusal kombinasyonlarla sınırlanmadı. Veride daha karmaşık bir düzenlilik saptarsa, bu örüntüyü öngörülerini iyileştirmek için kullanabilirdi.

Model bir kefalet/kefalet kararı yok cevabından çok, bir puanla sayısallaştırılmış bir kaçma riski tahmini üretecek şekilde tasarlandı. Bu yaklaşım, kabul edilebilir maksimum risk eşiğinin, yani üstüne çıkıldığında sanığa kefalet hakkı verilmemesi gereken risk seviyesinin bir modelin yapamayacağı değerlendirici bir yargı gerektirdiğini doğrular. Bununla birlikte araştırmacılar risk eşiği nerede sabitlenirse sabitlensin, modellerinin kestirici puanını kullanmanın insan karar vericilerin performansının üstünde bir iyileşmeyle sonuçlanacağını hesapladılar. Mullainathan'ın ekibi, risk eşiğinin kefalet hakkı tanınmayan kişi sayısı kararı yargıçların verdiği zamanla aynı olacak şekilde ayarlanması hâlinde, suç oranlarının %24'e kadar azaltılabileceğini çünkü yeniden suç işleme olasılığı en yüksek olan sanıkların parmaklıkların arkasında olacağını hesapladılar. Araştırmacılar, tam tersi bir şekilde risk eşiği, kefalet hakkı tanınmayan insanların sayısı suç artmadan en aza indirilecek şekilde ayarlanırsa, alıkonan insan sayısının %42'ye kadar azaltılabileceğini de hesapladılar. Başka bir deyişle bir yapay öğrenme algoritmasını bu verilerle eğittiler. Söz konusu

algoritma hangi sanıkların yüksek riskli olduğunu öngörmek konusunda insan yargıçlardan daha iyi sonuç alıyordu.

Yapay öğrenmeyle inşa edilen model ayrıca aynı bilgiyi kullanan doğrusal modellerden çok daha başarılıydı. Nedeni ise ilgi çekiciydi: "Yapay öğrenme algoritmaları başka şekilde atlanabilecek değişken kombinasyonlarında önemli sinyalleri bulur." Algoritmanın başka yöntemlerin kolayca kaçırdığı örüntüleri bulma becerisi algoritmanın en yüksek riskli olarak sınıflandırdığı sanıklar için özellikle daha belirgindir. Başka bir deyişle verideki bazı örüntüler nadiren olmakla birlikte güçlü bir şekilde yüksek riski öngörür. Bu bulgu –algoritmanın nadir ama kesin örüntüleri yakalaması– bizi tekrar kırık bacaklar kavramına getiriyor.

Araştırmacılar algoritmayı ayrıca her yargıcın, Bölüm 9'da tarif ettiğimiz yargıç modeline benzer bir modelini inşa etmek için kullandılar (ancak basit doğrusal kombinasyonlarla kısıtlamadan). Bu modelleri bütün veri setine uygulamak takıma, aynı vakalara yargıçların verecekleri kararları simüle etme ve kararları karşılaştırma imkânı verdi. Sonuçlar kefalet davalarında hatırı sayılır miktarda sistem gürültüsünü işaret etti: Bu gürültünün bir kısmı yatay gürültü idi: Yargıçlar yumuşaklıklarına göre ayrıldıklarında, en yumuşak beşte bir (yani yargıçların en yüksek kefaletle salma oranına sahip %20'lik kısmı) sanıkların %83'ünü serbest bırakırken, en az yumuşak beşte bir sadece %61'ini kefaletle salıyordu. Yargıçlar ayrıca hangi sanıkların yüksek kaçış riskine sahip olduğu konusunda çok farklı yargı örüntüleri sergiliyordu. Bir yargıç tarafından düşük kaçma riskli olarak görülen bir sanık, genelde daha katı olmayan bir başka yargıç tarafından yüksek kaçma riskli olarak görülebiliyordu. Bu sonuçlar çok net örüntü gürültüsü kanıtı sunuyordu. Daha detaylı bir analiz varyansın %67'sinden vakalar arasındaki farkların, %33'ünden de sistem gürültüsünün sorumlu olduğunu ortaya koydu. Sistem gürültüsü bir

miktar yatay gürültü –ortalama katılıktaki farklar– de içeriyordu ama büyük kısmı (%79'u) örüntü gürültüsüydü.

Son olarak ve neyse ki, yapay öğrenme programının ürettiği daha büyük isabetlilik, başta ırksal adalet olmak üzere yargıçların güttüğü diğer teşhis edilebilir hedefler harcanarak sağlanmaz. Teorik olarak algoritma hiç ırk verisi kullanmasa da program ırksal farklılıkları kasıtsız olarak ağırlaştırabilir. Bu farklılıklar modelin ırkla fazlasıyla bağıntılı (posta kodu gibi) kestiriciler kullanması durumunda ya da algoritmanın eğitildiği veri kaynağı ön yargılı olduğunda doğabilir. Örneğin geçmişteki tutuklanma sayısı bir kestirici olarak kullanılırsa ve eğer geçmiş tutuklanmalar ırk ayrımcılığından etkilenmişse, o zaman bundan doğan algoritma da ayrım yapacaktır.

Bu tür bir ayrımcılık prensipte elbette bir risk olmakla birlikte, bu algoritmanın kararları önemli açılardan yargıçlarınkinden ırk yönünden daha çok değil, daha az ön yargılıdır. Örneğin, risk eşiği yargıçların kararlarıyla aynı suç oranına ulaşacak şekilde sabitlendiyse, o zaman algoritma %41 daha az renkli insanı hapse atar. Benzer sonuçlar başka senaryolarda da bulunur: Doğruluk açısından kazançlar ırksal farklılıkları illa alevlendirmez ve yine araştırma ekibinin gösterdiği gibi algoritma kolayca bu farklılıkları azaltacak şekilde yönlendirilebilir.

Farklı bir alandaki başka bir çalışma, algoritmaların nasıl eş zamanlı olarak doğruluğu artırıp ayrımcılığı azaltabileceğini gösteriyor. Columbia Business School'dan Bo Cowgill büyük bir teknoloji şirketinde yazılım mühendislerinin işe alınmasını inceledi. Mülakata girecek adayları seçmek için, insan öz geçmiş eleyicileri kullanmak yerine, Cowgill adayların öz geçmişlerini elemek için yapay öğrenen bir algoritma geliştirdi ve algoritmayı şirketin aldığı ve değerlendirdiği üç yüz binden fazla başvuru üzerinden eğitti. Algoritmanın seçtiği adayların mülakatların ardından bir iş teklifi alma olasılığı insanlar tarafından seçilen

adaylardan %14 daha yüksekti. Adaylar teklif aldıkları zaman, algoritma grubundakilerin teklifi kabul etme ihtimali insanlar tarafından seçilen gruptakilerden %18 daha yüksekti. Algoritma ayrıca ırk, cinsiyet ve diğer ölçüler açısından, daha çeşitli bir aday grubu seçmişti. Elit bir okuldan mezun olmamış, iş tecrübesi daha az ve bir referansı olmayan gibi, "geleneksel olmayan" adaylar seçme olasılığı daha yüksekti. İnsan karar alıcılar "tipik" yazılım mühendisi profilinin bütün özelliklerini karşılayan öz geçmişleri tercih etme eğilimindeydiler ama algoritma ilgili her kestiriciye uygun ağırlığı vermişti.

Açık olmak gerekirse bu örnekler, algoritmaların her zaman adil, ön yargısız ve ayrımcılıktan arınmış olduğunu ispatlamıyor. Tanıdık bir örnek, iş adaylarının başarısını öngörmesi gereken ancak aslında geçmiş terfi kararlarının bir örneği üzerinden eğitilen bir algoritma olabilir. Elbette bu tür bir algoritma geçmiş terfi kararlarındaki bütün insani ön yargıları tekrarlayacaktır.

Irk ve cinsiyet farklılıklarını sürdüren bir algoritma inşa etmek mümkün ve belki de çok kolaydır ve bunu yapan birçok algoritma örneği bildirilmiştir. Bu örneklerin görünürlüğü algoritmaya dayalı karar alımındaki ön yargı konusunda endişelerin artmasını açıklar. Bununla birlikte, algoritmalar hakkında genel sonuçlara varmadan önce, bazı algoritmaların insan yargıçlardan sadece daha doğru değil, aynı zamanda daha adil olduklarını hatırlamamızda fayda var.

Kuralları Neden Daha Sık Kullanmayız?

Bu kısa mekanik karar alma turunu özetlemek için, her tür kuralın insan yargısına üstünlüğünün iki nedenini gözden geçirelim. Öncelikle Bölüm 9'da anlatıldığı gibi, sadece en yeni ve en karmaşık olanlar değil, bütün mekanik öngörü teknikleri insan yargısına göre önemli ilerlemeleri temsil etmektedir. Kişisel örüntülerin ve

durum gürültüsünün kombinasyonu insan yargılarının kalitesi üstünde öyle büyük bir ağırlık oluşturur ki, basitlik ve gürültüsüzlük oldukça büyük avantajlardır. Sadece akla uygun olan basit kurallar bile tipik olarak insan yargısından daha iyi iş çıkarırlar.

İkinci olarak, bazen veriler sofistike YZ tekniklerinin geçerli örüntüleri saptayacağı ve basit bir modelin kestirici gücünün ötesine geçebileceği kadar zengindir. YZ bu şekilde başarılı olduğunda, bu modellerin insan yargısına avantajı sadece gürültünün yokluğuyla sınırlı kalmaz, çok daha fazla bilgiden istifade etme becerisi de eklenir.

Bu avantajlar ve onları destekleyen kanıtların büyüklüğü göz önüne alındığında, algoritmaların bu kitapta ele aldığımız profesyonel yargı türleri için neden daha kapsamlı bir şekilde kullanılmadığını sormak yerinde olacaktır. Algoritmalar ve yapay öğrenmeyle ilgili bütün heyecanlı konuşmalara ve belli alanlardaki önemli istisnalara rağmen, bunların kullanımları hâlâ sınırlıdır. Birçok uzman kendi yargılarına güvenmeyi yeğleyerek "mekaniğe karşı klinik" tartışmasını yok saymaktadırlar. Kendi sezgilerine güvenir ve makinelerin daha iyi iş çıkaracağından şüphe duyarlar. Algoritmaya dayalı karar alma fikrini makineleşmek ve sorumluluklarından mahrum bırakılmak olarak görürler.

Örneğin tıbbi teşhislerde algoritmaların kullanılması etkileyici ilerlemelere rağmen, henüz rutinleşmedi. Çok az örgüt işe alım ve terfi kararlarında algoritma kullanıyor. Hollywood stüdyolarının üst düzey yöneticileri filmlere yeşil ışığı bir formüle değil, yargılarına ve tecrübelerine dayanarak yakıyorlar. Kitap yayıncıları da öyle. Dahası Michael Lewis'in çok satan kitabı *Moneyball*'da söylendiği gibi istatistik takıntılı Oakland Athletics beyzbol takımının hikâyesi böyle bir iz bıraktıysa, nedeni algoritmaya dayalı kesinliğin spor takımlarının karar alma süreçlerinde çok uzun süre kural değil, istisna olmasıdır. Bugün bile koçlar, menajerler ve onlarla çalışan

insanlar sıklıkla kendi iç seslerine güveniyor ve istatistiki analizin iyi yargının yerini alamayacağında ısrar ediyorlar.

Meehl ve bir diğer yazar 1996 yılında birlikte kaleme aldıkları bir makalede, psikiyatristlerin, doktorların, yargıçların ve diğer profesyonellerin mekanik yargıya tam on yedi tür itirazını listelediler (ve çürüttüler). Yazarlar klinik tedavi uzmanlarının direncinin "teknolojik işsizlik korkusu", "kötü eğitim" ve "genel bir bilgisayar karşıtlığının" da aralarında olduğu sosyopsikolojik faktörlerin bir kombinasyonuyla açıklanabileceği sonucuna vardılar.

O zamandan beri, araştırmacılar bu dirence katkı sağlayan ek faktörleri saptıyorlar. Burada bu araştırmanın tam bir incelemesini sunmayı hedeflemiyoruz. Bu kitapta amacımız, Yargıç Frankel'in ifade edeceği gibi, "insanların yerini makinelerin almasını" savunmak değil, insan yargısının iyileşmesi için önerilerde bulunmak.

Ancak insanları mekanik öngörülere direnmeye nelerin ittiği konusundaki bazı bulgular insan yargısını tartışmamız açısından önem taşıyor. Yakın zamanda yapılan araştırmalardan kilit bir içgörü doğdu: İnsanlar algoritmalardan sistematik şekilde şüphe duymuyorlar. Örneğin bir insandan ve bir algoritmadan tavsiye almak arasında seçim yapma şansı sunulduğunda, sıklıkla algoritmayı yeğliyorlar. Algoritmalara direnç ya da *algoritmadan kaçınma,* kendini her zaman yeni karar destek araçları benimsemeyi kapsamlı bir reddediş şeklinde göstermiyor. Daha çok, insanlar bir algoritmaya şans vermeye istekli oluyor ama hatalar yaptığını görür görmez ona güvenmeyi bırakıyorlar.

Bir düzeyde bu tepki akla yatkın görünüyor: Güvenemeyeceğiniz bir algoritmayla neden uğraşasınız ki? İnsanlar olarak hatalar yaptığımızın fazlasıyla farkındayız ama bu paylaşmaya hazır olmadığımız bir ayrıcalık. Makinelerin kusursuz olmasını bekliyoruz. Bu beklenti çiğnenirse, onları bertaraf ediyoruz.

Öte yandan insanlar bu sezgisel beklenti nedeniyle bu tercih kanıtlanabilir şekilde daha alt seviye sonuçlar ürettiğinde bile,

algoritmalara güvenmeme ve kendi yargılarını kullanma eğilimindeler. Bu tavrın kökleri derinlere uzanıyor ve kusursuza yakın kestirimci doğruluğa erişilmediği sürece değişeceğe benzemiyor.

Neyse ki kuralları ve algoritmaları daha iyi yapan şeylerin büyük kısmı insan yargısında kopyalanabiliyor. Bilgiyi bir YZ modeli kadar etkin kullanmayı umamayız ama basit modellerin basitliğine ve gürültüsüzlüğüne öykünebiliriz. Sistem gürültüsünü azaltan yöntemler benimsediğimiz ölçüde kestirimci yargıların kalitesinde iyileşmeler görebiliriz. Yargılarımızı nasıl iyileştireceğimiz konusu ise Kısım 5'in ana temasını oluşturuyor.

Kurallar ve Algoritmalardan Bahsederken

"Çok fazla veri olduğunda, yapay öğrenme algoritmaları insanlardan ve basit modellerden daha iyi iş çıkarır. Ancak en basit kurallar ve algoritmalar bile insan yargıçlar karşı çok büyük avantajlara sahiptir; gürültüden arınmışlardır ve karmaşık ve kestiriciler konusunda genellikle geçersiz içgörüler uygulamaya çalışmazlar."

"Tahminde bulunacağımız sonuç konusunda veri eksiğimiz olduğuna göre, neden bir eşit ağırlık modeli kullanmıyoruz? Neredeyse uygun bir model kadar iyi iş çıkaracaktır ve vakaları tek tek ele alan insan yargısından daha iyi sonuç vereceği kesindir."

"Modelin öngörüsüne itirazınız var. Anlıyorum. Ama ortada bir kırık bacak mı var, yoksa sadece tahmin mi hoşunuza gitmedi?"

"Algoritma da hata yapar elbette. Ama ya insan yargıçlar daha fazla hata yapıyorsa, o zaman kime güvenmeliyiz?"

BÖLÜM 11

Objektif Bilgisizlik

Son iki bölümün materyalini, insan yargısının kısıtlı başarısıyla ilgili uyandırıcı bulgularıyla birlikte, üst düzey yöneticilerden oluşan dinleyici gruplarıyla paylaşma tecrübemiz oldu. Aktarmayı hedeflediğimiz mesaj yarım yüzyıldan uzun süredir orada ve az sayıda karar alıcının buna maruz kalmaktan kaçındığını düşünüyoruz. Ama direnmeyi başardıkları kesin.

İzleyici gruplarımızdaki yöneticilerden bazıları bize gururla kendi iç seslerine bütün analizlerden daha çok güvendiklerini söylediler. Daha pek çokları da bu kadar dobra olmasa da aynı görüşü paylaşıyorlar. Yönetimsel karar alma konulu araştırmalar üst düzey yöneticilerin, özellikle daha kıdemli ve tecrübeli olanların, çeşitli ifadelerle *sezgi*, *iç ses* ya da basitçe *yargı* (bizim bu kitapta kullandığımızdan farklı bir anlamda) denen bir şeye bolca sığındıklarını gösteriyor.

Kısaca karar alıcılar, kendi iç seslerini dinlerler ve çoğu duyduklarından hoşnut görünür. Bu da akla şu soruyu getiriyor: Otorite ve büyük öz güven kombinasyonuyla kutsanmış bu insanlar iç seslerinden tam olarak ne duyuyorlar?

Yönetimsel karar almada sezgiyle ilgili bir çalışma, onu "belli bir hareket tarzı için bir aura, güçlü doğruluk kanaati ya da akla yatkınlıkla ama açıkça ifade bulan neden ve gerekçeler olmadan, esasen 'bilerek' ama nedenini bilmeden akla gelen bir yargı" olarak tanımlıyor. Nedenini bilmeden bilme duygusunun aslında Bölüm 4'te bahsettiğimiz yargının tamamlandığına dair *içsel sinyal* olduğunu düşünüyoruz.

İçsel sinyal kişinin kendine verdiği, bir yargıda kapanışa ulaşırken insanların elde etmek için çok çabaladığı (bazen çok fazla da değil) bir ödüldür. Değerlendirilen kanıtın ve ulaşılan yargının doğru geldiği tatmin edici bir duygusal tecrübe, hoşa giden bir tutarlılık hissidir. Yapbozun bütün parçaları oturur gibi gelir (Daha sonra, bu tutarlılık hissinin sık sık uymayan kanıt parçaları saklanarak ya da yok sayılarak güçlendiğini göreceğiz).

Bu içsel sinyali önemli –ve yanlışa yönlendirici– kılan, bir his olarak değil, bir inanç olarak yorumlanmasıdır. Bu duygusal tecrübe ("kanıt doğru geliyor") kişinin yargısının geçerliliğinde akılcı bir güven kılığına girer ("Nedenini bilmesem de biliyorum.").

Ancak güven bir isabetlilik garantisi değildir ve çoğu öz güvenli öngörü yanlış çıkar. Ön yargı ve gürültü öngörü hatalarına neden olurken, bu tür hataların en büyük kaynağı kestirimci yargıların ne kadar iyi *olduğunun* sınırı değildir. Ne kadar iyi *olabileceklerinin* sınırıdır. *Objektif bilgisizlik* dediğimiz bu sınır, bu bölümün odağı olacak.

Objektif Bilgisizlik

Kendinizi tekrarlı kestirimci yargılara varırken bulduğunuzda, kendinize sorabileceğiniz bir soruyu inceleyelim. Soracağınız bu soru, her göreve, örneğin hisse senedi seçmeye ya da profesyonel sporcuların performansını öngörmeye uygulanabilir. Ama basitlik yararına, Bölüm 9'da kullandığımız örneği seçeceğiz: iş adayla-

rının seçimi. Seneler içinde yüz adayı değerlendirdiğinizi düşünün. Yaptığınız değerlendirmeleri adayların o zamandan bu yana objektif olarak değerlendirilmiş performanslarıyla karşılaştırarak kararlarınızın ne kadar iyi olduğunu değerlendirme şansınız var. Rastgele bir çift adayı seçtiğinizde, en baştaki yargınız ve sonraki değerlendirmeleriniz hangi sıklıkla örtüşüyor? Başka bir deyişle herhangi iki adayı karşılaştırdığınızda, daha fazla potansiyele sahip olduğunu düşündüğünüz adayın, gerçekten daha yüksek performans sergileme olasılığı nedir?

Üst düzey yönetici gruplarıyla bu soru aracılığıyla sık sık gayriresmî anketler yaparız. En sık aldığımız cevaplar %75-%85 aralığındadır ve bu tepkilerin alçak gönüllülük ve övünür gibi görünmeme isteğiyle kısıtlandığından şüphe duyuyoruz. Özel, bire bir görüşmeler gerçek güven hissinin sıklıkla daha yüksek olduğunu ortaya koyuyor.

Artık yüzdelik uyum istatistiğine aşina olduğunuz için, bu değerlendirmenin doğurduğu sorunu kolayca görebiliyorsunuzdur. %80'lik bir YU kabaca 0,80'lik bir korelasyona karşılık gelir. Bu kestirici güç seviyesine gerçek dünyada nadiren ulaşılır. Personel seçimi alanında, yakın zamanda yapılan bir çalışma insan yargıçların performansının bu sayıya yaklaşmadığını ortaya koydu. Ortalama olarak, 0,28'lik bir kestirimci korelasyona (YU = %59) ulaşabiliyorlar.

Personel seçiminin zorluğunu düşündüğünüzde, hayal kırıklığına uğratan bu sonuçlar o kadar şaşırtıcı değil. Bugün yeni bir işe başlayan bir insan pek çok zorluk ve fırsatlarla karşılaşacaktır ve şans müdahil olarak hayatının yönünü pek çok şekilde değiştirecektir. Ona inanan, fırsatlar yaratan, onu terfi ettiren ve öz güven ve motivasyonunu artıran bir şefle karşılaşabilir. Hiçbir kabahati olmadan, şansı daha az yaver gittiği için, kariyerine moral bozucu bir başarısızlıkla başlayabilir. Kişisel hayatında da iş performansını etkileyen olaylar olabilir. Bu olayların ve şartların hiçbiri bugün ne siz ne başkası ne de dünyanın en iyi kestirimci modeli tarafından

öngörülebilir. Ele avuca sığmaz belirsizlik öngörmeye çalıştığınız sonuç hakkında şu anda bilinemeyecek her şeyi kapsar.

Dahası adaylar hakkındaki çok şey prensipte bilinebilir ama siz kararınızı verirken bilinmiyordur. Bizim amaçlarımız açısından, bu bilgi boşluklarının yeterli kestirimci test eksikliğinden mi, sizin daha fazla bilgi edinmenin maliyetinin haklı gerekçesi olmadığı kararınızdan mı yoksa olgu bulma ihmalinizden mi kaynaklandığının bir önemi yok. Öyle ya da böyle, kusursuza uzak bir bilgi evresindesiniz.

Ele avuca sığmaz belirsizlik (bilinmesi mümkün olmayan) ve kusurlu bilgi (bilinebilen ama bilinmeyen) kusursuz öngörüyü imkânsız kılar. Bu bilinmeyenler kararınızda ön yargı ya da gürültü sorunları değildir; görevin objektif özellikleridir. Önemli bilinmeyenlere dair bu objektif bilgisizlik erişilebilir doğruluğu ciddi anlamda kısıtlar. Burada terminoloji açısından özgürlüğümüzü ilan edip yaygın olarak kullanılan *belirsizlik* sözcüğünü *bilgisizlik* ile değiştiriyoruz. *Bilgisizlik* terimi dünyayla ve gelecekle ilgili olan belirsizlikle, aynı olması gereken kararlardaki farklılığı açıklayan gürültü arasında karışıklık yaşanması riskini sınırlandırır.

Bazı durumlarda diğerlerinden daha fazla bilgi (ve daha az objektif bilgisizlik) olur. Çoğu profesyonel yargı hayli iyidir. Birçok hastalıkta doktorların öngörüleri kusursuzdur ve pek çok hukuki anlaşmazlıkta, avukatlar yargıçların muhtemelen ne karar vereceğini size büyük bir doğruluk payıyla söyleyebilir.

Ancak genel olarak kestirici görevlerle uğraşan insanların objektif bilgisizliklerini hafife almalarını rahatça bekleyebilirsiniz. Aşırı öz güven en sık kanıtlanan bilişsel ön yargılardan biridir. Özellikle kişinin sınırlı bilgiyle bile kesin tahminlerde bulunma becerisiyle ilgili yargılar, herkesçe bilinecek kadar aşırı öz güvenlidir. Kestirici yargılarda gürültü hakkında söylediklerimiz objektif bilgisizlik için de söylenebilir: Tahminin olduğu yerde bilgisizlik vardır, sandığınızdan çok daha fazla.

Aşırı Öz Güvenli Üstatlar

İyi bir dostumuz olan psikolog Philip Tetlock, hakikate sıkı sıkıya bağlıdır ve muzip bir espri anlayışına sahiptir. 2005 yılında, *Expert Political Judgment* adında bir kitap yayımladı. Nötr bir his veren başlığa rağmen, kitap uzmanların politik olaylar hakkında doğru tahminlerde bulunma becerilerine yıkıcı bir saldırı niteliğindeydi.

Tetlock yaklaşık üç yüz uzmanın öngörülerini inceledi: önde gelen gazeteciler, saygın akademisyenler ve ulusal liderlerin üst düzey danışmanları. Politik, ekonomik ve sosyal öngörülerinin gerçekleşip gerçekleşmediğini sordu. Araştırma yirmi yıla yayılıyordu; uzun vadeli tahminlerin doğru olup olmadığını öğrenmek için sabırlı olmak gerekiyordu.

Tetlock'un kilit bulgusu, sözde uzmanların büyük ve önemli politik hadiselerdeki tahminlerinde insanı sersemletecek kadar başarısız olmasıydı. Kitap akılda yer eden şu cümlesiyle ün kazandı: "Ortalama uzman kabaca dart atan bir şempanze kadar isabetliydi." Kitabın daha kesin bir diğer beyanı ise hayatlarını "politik ve ekonomik trendler hakkında yorum yaparak ya da tavsiyelerde bulunarak" kazanan uzmanların, "gazetecilerin ve *New York* Times'ın dikkatli okurlarının ortaya çıkan durumları 'okumada' olduğundan daha iyi" olmadıklarıydı. Elbette uzmanlar harika hikâyeler anlatıyordu. Bir durumu analiz edebiliyor, nasıl ilerleyeceği konusunda sürükleyici bir tablo çizebiliyor ve onlara katılmayan kişilerin itirazlarını televizyon stüdyolarında büyük bir öz güvenle çürütüyorlardı. Ama aslında neler olacağını biliyorlar mıydı? Pek sayılmaz.

Tetlock bu sonuca hikâye anlatımı kısmını atlayarak ulaştı. Her konu için uzmanlardan üç olası sonuca, –statüko, bir şeyden daha fazlası ya da daha azı– olasılık tayin etmelerini istedi. Dart atan bir şempanze bu sonuçlardan herhangi birini, gerçeklikten tamamen bağımsız olarak, aynı olasılıkla –üçte bir– "seçerdi". Tetlock'un uzmanları bu çok düşük standardı güçlükle geçebildi-

ler. Ortalama olarak, gerçekleşen olaylara gerçekleşmeyenlerden birazcık daha yüksek olasılık tanıdılar ama performanslarının en çarpıcı özelliği öngörülerine aşırı güvenleriydi. Dünyanın nasıl işlediği konusunda net teorilerle kutsanmış üstatlar en öz güvenli ama en az isabetli olanlardı.

Tetlock'un bulguları belli olaylarda detaylı uzun vade tahminlerinin basitçe imkânsız olduğunu ortaya koyuyordu. Dünya en küçük olayların bile büyük sonuçlara yol açabildiği karmaşık bir yer. Örneğin, ana rahmine düşmede, tarihteki bütün önemli şahsiyetlerin (ve önemsizlerin) farklı bir cinsiyetle doğma ihtimali olduğunu düşünün. Öngörülemez olaylar gerçekleşebilir ve bu öngörülemez olayların sonuçları da öngörülemezdir. Sonuç olarak siz gelecekte daha uzağa baktıkça objektif bilgisizlik aralıksız olarak birikir. Uzman politik yargının sınırı öngörücülerin bilişsel sınırıyla değil, geleceğe dair ele avuca sığmaz objektif bilgisizlikleriyle belirlenir.

Bu durumda vardığımız sonuç, üstatların uzak öngörülerinin başarısızlığı için suçlanmaması gerektiğidir. Bununla birlikte imkânsız bir göreve yeltendikleri ve başarılı olabileceklerine inandıkları için bir miktar eleştiriyi hak ediyorlar.

Tetlock şok edici çok uzun vadeli öngörülerin işe yaramadığı keşfinden seneler sonra, eşi Barbara Mellers'la güçlerini birleştirerek insanların nispeten kısa vadede –genellikle bir yıldan kısa sürede– dünya olayları hakkında öngörüde bulunmaları istendiğinde ne kadar iyi çıkardığını araştırmaya soyundu. Ekip, kısa vadeli tahminin zor olduğunu ama imkânsız olmadığını ve Tetlock ve Mellers'ın *süper tahminciler* olarak adlandırdıkları bazı kişilerin istihbarat topluluğundaki profesyonellerin de aralarında olduğu diğerlerinden şaşmaz bir şekilde daha iyi olduğunu keşfetti. Burada kullandığımız terimlerle, bu yeni bulguları objektif bilgisizliğin gelecekte daha ileri bakıldığında arttığı kavramıyla tutarlıydı. Süper tahmincilere Bölüm 21'de geri döneceğiz.

Yetersiz Yargıçlar ve Sadece Biraz Daha İyi Modeller

Tetlock'un ilk araştırması insanların genel olarak uzun vadeli politik öngörülerde iyi iş çıkaramadıklarını göstermişti. Net bir kristal küresi olan tek bir insan bulmak bile sonuçları tamamen değiştirirdi. Bir görev ancak çok sayıda güvenilir aktörün şanslarını deneyip başarısız olmasının ardından imkânsız sayılabilir. Bilginin mekanik birleştirilmesinin genellikle insan yargısına üstün geldiğini gösterdiğimiz gibi, kuralların ve algoritmaların kestirici doğruluğu da sonuçların özünde ne kadar öngörülebilir ya da öngörülemez olduğu konusunda daha iyi bir test sağlar.

Önceki bölümler sizde algoritmaların kestirici yargılardan ezici şekilde üstün olduğu izlenimini yaratmış olabilir. Modeller tutarlı olarak insanlardan daha iyiler ama çok daha iyi değiller. Aynı bilgiyle insanların kötü, modellerin çok iyi iş çıkardığı durumlara dair herhangi bir kanıt yok.

Bölüm 9'da mekanik birleştirmenin klinik yargıya üstünlüğünü gösteren 136 çalışmanın gözden geçirilmesinden bahsettik. O üstünlüğün kanıtı gerçekten "büyük ve tutarlı" olmakla birlikte, performans farkı çok büyük değil. İncelemedeki çalışmaların doksan üç tanesi ikili kararlara odaklanmış ve klinik tedavi uzmanlarının ve formüllerin "isabet oranı"nı ölçmüştü. Medyan çalışmada klinik tedavi uzmanları vakaların %68'inde, formüller ise %73'ünde haklı çıkmıştı. 35 çalışmadan oluşan daha küçük bir alt kümede korelasyon katsayısı bir doğruluk ölçüsü olarak kullanıldı. Bu çalışmalarda, klinik tedavi uzmanları 0,32'lik bir sonuçla (YU = %60) medyan bir korelasyona ulaşırken, formüller 0,56'lık bir korelasyon (YU = %69) yakaladılar. Her iki ölçüde de formüller tutarlı olarak klinik tedavi uzmanlarından daha iyi sonuç çıkardılar ama mekanik kestirimlerin sınırlı geçerliliği çarpıcılığını koruyordu. Modellerin performansı nispeten alçak bir öngörülebilme tavanının tablosunu değiştirmiyordu.

Ya yapay zekâ? Belirttiğimiz gibi, YZ genellikle basit modellerden daha iyi performans sağlar. Ancak pek çok uygulamada, performansı kusursuz olmaya çok uzaktır. Örneğin, Bölüm 10'da tartıştığımız kefalet öngörü algoritmasını düşünün. Algoritmanın kefalet hakkı tanınmayan insanların sayısını sabit tutarak suç oranlarını %24'e kadar azaltabildiğini belirtmiştik. Bu insan kefalet yargıçlarının öngörülerine göre etkileyici bir iyileşme ancak algoritma hangi sanıkların yeniden suç işleyeceklerini kusursuz bir isabetle öngörebilseydi, suç oranını çok daha fazla azaltabilirdi. *Azınlık Raporu* filminde gelecekte işlenecek suçların doğaüstü öngörülerinin bilim kurgu olmasının bir nedeni var: İnsan davranışını öngörmek çok büyük miktarda objektif bilgisizlik içerir.

Sendhil Mullainathan ve Ziad Obermeyer'in yürüttüğü bir başka çalışma, kalp krizi teşhislerini modelledi. Hastalar olası bir kalp krizi işaretleri sergilediklerinde, acil servis doktorlarının ek testler isteyip istememek konusunda karar vermeleri gerekir. Prensipte, hastalar sadece kalp krizi riski yeterince yüksek olduğunda test edilmelidir çünkü testler sadece maliyetli değil, aynı zamanda kullanışsız ve risklidir; düşük riskli hastalar için istenen bir durum değildir. Dolayısıyla bir hekimin test isteme kararı kalp krizi riskinin değerlendirilmesini gerektirir. Araştırmacılar bu değerlendirmeyi yapmak için bir YZ modeli oluşturdular. Model iki bin dört yüz değişken kullanıyor ve büyük bir vaka örneğine dayanıyordu (1,6 milyon hastanın 4,4 milyon federal sağlık sigortası programı incelemesi). Bu miktarda veriyle, model muhtemelen objektif bilgisizlik sınırına yaklaşmıştı.

Şaşırtıcı olmayan bir sonuçla, YZ modelinin doğruluğu doktorlarınkinden bariz olarak üstün çıktı. Modelin performansını değerlendirmek için, modelin risk açısından en yüksek onda birlik dilime yerleştirdiği hastaları düşünün. Bu hastalar teste tabi tutulduklarında, %30'unun kalp krizi geçirmiş olduğu ve risk dağılımının ortasındaki hastaların %9,3'ünün bir kalp krizi yaşadığı anlaşıldı.

Bu ayrım seviyesi etkileyici olmakla birlikte kusursuz olmaya çok uzaktır. Makul bir şekilde, doktorların performansının, en az yargılarının kusurları kadar objektif bilgisizliğin kısıtlamalarıyla da sınırlandığı sonucuna varabiliriz.

Bilgisizliğin İnkârı

Kusursuz öngörünün imkânsızlığında ısrar ederken bariz bir olguyu ifade eder gibi görünüyor olabiliriz. İtiraf etmek gerekir ki geleceğin öngörülemez olduğunu iddia etmek büyük bir kavramsal buluş sayılmaz. Bununla birlikte kestirimci aşırı öz güvenle ilgili tutarlı bulguların gösterdiği gibi, olgunun aşikârlığı yok sayılma sıklığıyla yarışır.

Aşırı öz güvenin baskınlığı iç sese güvenen karar alıcılarla ilgili gayriresmî anketimize yeni bir ışık tutuyor. İnsanların öznel öz güven hislerini sık sık kestirimci geçerlilik göstergesiyle karıştırdıklarından bahsetmiştik. Örneğin Bölüm 9'da Nathalie ve Monica'yla ilgili kanıtları gözden geçirdikten sonra, tutarlı bir yargıya vardığınız zaman hissettiğiniz içsel sinyal, size adaylardan Nathalie'nin daha güçlü olduğu güvenini verdi. Bununla birlikte, bu öngörüden emin olduysanız, geçerlilik yanılsamasına kanmışsınız demektir zira size verilen bilgiyle erişebileceğiniz doğruluk hayli düşüktü.

Olağanüstü yüksek düzeyde kestirimci doğruluğa ulaşabileceklerine inanan insanlar sadece aşırı öz güvenli değildir. Sadece yargılarındaki gürültü ve ön yargı risklerini inkâr etmekle kalmazlar. Ya da sadece kendilerini diğer ölümlülerden daha üstün görmezler. Aslında öngörülemez olan olayların öngörülebilirliğine de inanarak üstü kapalı bir şekilde belirsizlik gerçeğini inkâr ederler. Burada kullandığımız terimlerle, bu tavır *bilgisizliğin inkârı* demektir.

Bilgisizliğin inkârı Meehl ve takipçilerini şaşırtan bulmacaya, Meehl'in mesajının neden kale alınmadığı ve karar alıcıların neden

kendi sezgilerine bel bağlamaya devam ettiği sorusuna da bir cevap ekliyor. Karar alıcılar kendi iç seslerine kulak verdikleri zaman, içsel sinyali duyar ve getirdiği duygusal ödülü hissederler. İyi bir karara ulaşıldığı iç sinyali, öz güvenin, "nedenini bilmeden bilmenin" sesidir. Ancak kanıtın gerçek kestirimci gücünün objektif olarak değerlendirilmesi bu güven seviyesini nadiren haklı çıkarır.

Sezgisel eminliğin duygusal ödülünden vazgeçmek kolay değildir. Etkileyici bir şekilde, liderler sezgisel karar almaya fazlasıyla belirsiz olarak algıladıkları durumlarda özellikle daha yatkın olduklarını söylerler. Olgular onlara çok istedikleri anlayış hissini ve öz güveni sağlamadığında, bunları bulmak için sezgilerine yönelirler. Bilgisizlik çok büyük olduğunda inkârı çok daha caziptir.

Bilgisizliğin inkârı bir başka bulmacayı daha açıklar: Burada sunduğumuz kanıtlarla karşı karşıya geldiklerinde, pek çok lider görünüşte çelişkili bir sonuç çıkarır. İç sese dayalı kararlarının kusursuz olmayabileceğini ama daha sistematik alternatifler de kusursuz olmaya uzaksa, benimsenmeye değmeyeceklerini söylerler. Örneğin insan yargıçlarla çalışan performansı arasındaki ortalama korelasyonun 0,28 (YU = %59) olduğunu hatırlayalım. Aynı çalışmaya göre ve incelediğimiz kanıtla tutarlı olarak, mekanik öngörü daha iyi iş çıkarabilir ama arada büyük fark olmaz: Kestirimci isabetliliği 0,44'tür (YU = %65). Bir üst düzey yönetici kendine, bu zahmete ne gerek var diye sorabilir.

Cevap, kimin işe alınacağıyla ilgili kararlar gibi önemli bir şeyde, geçerlilikteki bu artışın çok büyük bir değeri olduğudur. Aynı üst düzey yöneticiler bu kadar büyük olmayan kazançlar yakalamak için, çalışma tarzlarında rutin olarak önemli değişiklikler yaparlar. Mantıksal olarak başarının hiçbir zaman garantilenemeyeceğini ve kararlarında çabaladıkları şeyin daha yüksek bir başarı şansı olduğunu anlarlar. Ayrıca olasılığı da anlarlar. Aynı fiyata kazanma şansı %65 olan bir piyango bileti dururken, şansı %59 olan bir bileti hiçbiri satın almaz.

Buradaki zorluk, bu durumdaki "fiyatın" aynı olmamasıdır. Sezgisel yargı, ödülüyle, yani iç sinyalle birlikte gelir. İnsanlar daha yüksek bir isabet seviyesine ulaşacak bir algoritmaya güvenmeye hazırdırlar çünkü bu onlara iç sinyalin sağladığına denk ya da onu aşan bir kesinlik hissi sağlar. Ancak alternatifi yüksek geçerlilik iddiası bile olmayan bir tür mekanik süreç iken iç sinyalin duygusal ödülünden vazgeçmek yüksek bir bedeldir.

Bu gözlem, yargının iyileşmesi açısından önemli bir saklı anlam taşır. Mekanik ve algoritmaya dayalı öngörü yöntemlerinin lehine olan bütün kanıtlara ve kestirimci isabetlilikteki artımlı iyileşmenin değerini açıkça ortaya koyan rasyonel matematik hesabına rağmen, pek çok karar alıcı onları kendi sezgilerini kullanma becerisinden mahrum bırakacak karar alma yaklaşımlarını reddeder. Algoritmalar kusursuz olmaya yaklaşamadığı sürece –ve pek çok alanda objektif bilgisizlik asla yaklaşamayacaklarını söylüyor– insan yargısının yeri doldurulamaz. Bu yüzden iyileştirilmelidir.

Objektif Bilgisizlikten Bahsederken

"Öngörünün olduğu yerde bilgisizlik vardır, muhtemelen sandığımızdan da çok. Güvendiğimiz uzmanların dart atan şempanzelerden daha isabetli olup olmadıklarını kontrol ettik mi?"

"Gerçekten bildiğiniz bir şey yüzünden değil, bir iç sinyal yüzünden iç sesinize güvendiğinizde, objektif bilgisizliğinizi inkâr etmektesinizdir."

"Modeller insanlardan daha iyi iş çıkarır, ama arada büyük fark yoktur. Çoğunlukla vasat insan yargıları ve biraz daha iyi modellerle karşılaşırız. Yine de daha iyi iyidir ve modeller daha iyidir."

"Bu kararları almak için model kullanırken hiçbir zaman rahat olmamalıyız yeterli öz güvene sahip olmak için tek ihtiyacımız iç sinyaldir. Bu yüzden olası en iyi karar sürecine sahip olduğumuzdan emin olmalıyız."

BÖLÜM 12

Normal Vadisi

Şimdi daha geniş kapsamlı bir soruya dönüyoruz: Pek çok sorunun kolay olduğu ama daha pek çoğuna objektif bilgisizliğin hâkim olduğu bir dünyada konfora nasıl ulaşırız? Sonuçta objektif bilgisizliğin ciddi boyutta olduğu yerlerde, bir süre sonra insani işlerde kristal kürelerin işe yaramazlığının farkına varmalıyız. Ama dünyadaki alışılmış tecrübemiz bu değildir. Bunun yerine, bir önceki bölümün öne sürdüğü gibi, çok az miktarda işe yarar bilgiden gelecek hakkında cesur öngörülerde bulunmak için dizginsiz bir arzuyu koruruz. Bu bölümde, baskın ve yanılgılı bir yaklaşımla, öngörülemez olayların yine de anlaşılabileceği hissini ele alacağız.

Bu inanç aslında ne anlama gelir? Bu soruyu iki bağlamda soracağız. sosyal bilimin yürütülmesi ve günlük olayların tecrübesi.

Hayat Güzergâhını Öngörmek

2020 yılında, her ikisi de Princeton Üniversitesi'nde sosyoloji profesörü olan Sara McLanahan ve Matthew Salganik'in liderliğindeki 112 araştırmacı, *Proceedings of the National Academy*

of Sciences'ta sıra dışı bir makale yayımladılar. Araştırmacıların hedefi sosyal bilimcilerin sosyal açıdan kırılgan ailelerin hayat güzergâhlarında neler yaşanacağını ne kadar anladıklarını saptamaktı. Sosyal bilimciler ellerindeki bilgilerle, bir ailenin hayatında yaşanacak olayları ne kadar iyi öngörebilirlerdi? Bilhassa, sosyologların normalde topladığı ve araştırmalarına uyarladığı bilgileri kullanarak hayat olaylarını tahmin etmekte nasıl bir isabet düzeyi tutturabilirlerdi? Bizim terimlerimizle, çalışmanın amacı sosyologlar işlerini yaptıktan sonra bu hayat olaylarında kalan nesnel bilgisizlik seviyesini ölçmekti.

Yazarlar doğumdan on beş yaşına kadar takip edilen çocukların geniş ölçekli ve uzun vadeli bir incelemesi olan Kırılgan Aileler ve Çocuk Refahı Çalışması'ndan materyal çektiler. Büyük veri tabanı çoğu büyük Amerikan şehirlerinde evlenmemiş ebeveynlerden dünyaya gelmiş yaklaşık beş bin çocuğun aileleri hakkında binlerce bilgi maddesi içeriyor. Veriler çocukların büyükanne ve büyükbabalarının eğitim ve istihdam durumu, aile üyelerinin sağlığı hakkında detaylar, ekonomi ve sosyal statü endeksleri, çok sayıda anket cevabı, bilişsel kabiliyet ve kişilik testleri gibi konuları kapsıyor. Bu olağanüstü zenginlikte bir bilgi havuzu ve sosyal bilimciler bu bilgileri çok iyi kullandılar: Kırılgan Aileler Çalışması'nın verilerine dayanarak 750 bilimsel makale yazıldı. Bu makalelerin büyük kısmı lise notları ve sabıka kaydı gibi ileriki dönem sonuçlarını izah etmek için çocuklar ve aileleri hakkındaki arka plan verilerini kullandılar.

Princeton ekibinin yürüttüğü çalışma çocuk on beş yaşındayken gözlemlenen altı sonucun tahmin edilebilirliğine odaklandı. Bu sonuçlar arasında yakın dönemde bir ev tahliyesi yaşanıp yaşanmadığı, çocuğun not ortalaması ve hane halkının maddi şartlarının genel bir ölçümü de yer alıyordu. Araştırmayı organize edenler "ortak görev metodu" adını verdikleri şeyi kullandılar. Araştırma ekiplerini, Kırılgan Aileler Çalışması'nda yer alan her

aileyle ilgili erişilebilir verileri kullanarak seçili altı sonuç için isabetli tahminler üretme konusunda yarışmaya davet ettiler. Bu tür bir meydan okuma, sosyal bilimlerde yeni olmakla birlikte takımların standart bir metin grubunun makine tercümesi ya da büyük bir fotoğraf kümesinde bir hayvanın bulunması gibi görevlerde yarışmaya davet edildiği bilgisayar biliminde yaygındır. Bu tür yarışmalarda kazanan takımın başarısı zamanda bir noktada gelişme seviyesini tanımlar ve o seviye her zaman bir sonraki yarışmada aşılır. Hızlı hareketin beklenmediği bir sosyal bilim öngörü görevinde, yarışmada en isabetli tahmini bu verilerden alınacak sonucun tahmin edilebilirliğinin ölçüsü, başka bir deyişle objektif bilgisizliğin artık seviyesi olarak kullanmak akla yatkındır.

Bu meydan okuma araştırmacılar arasında hatırı sayılır bir ilgi uyandırdı. Son rapor çok daha büyük bir uluslararası başvuru havuzundan çekilen 160 hayli kalifiye ekipten sonuçları sundu. Seçilen yarışmacıların çoğu kendilerini veri bilimci olarak tanımlıyor ve yapay öğrenme kullanıyordu.

Yarışmanın birinci aşamasında katılımcı ekiplerin toplam örneğin yarısı için bütün veriye erişimi vardı ve bu veri altı sonucu da içeriyordu. Bu "eğitim verisini" kestirimci bir algoritma eğitmek için kullandılar. Algoritmaları daha sonra algoritmayı eğitmek için kullanılmayan ailelere ilişkin bir anlaşmazlık örneğinde kullanıldı. Araştırmacılar isabetliliği OKH kullanarak ölçtüler: Her örnek için öngörü hatası gerçek sonuçla algoritmanın öngörüsü arasındaki farkın karesiydi.

Kazanan modeller ne kadar iyiydi? Büyük bir veri setiyle eğitilen sofistike yapay öğrenme algoritmaları elbette, basit doğrusal modellerin tahminlerine üstünlük sağladılar (ve uzantı olarak insan yargılarından da daha iyi sonuç çıkaracaklardı). Ancak YZ modellerinin çok basit bir model üstüne getirdiği iyileşme çok küçüktü ve kestirimci doğrulukları hayal kırıklığı yaratacak ölçüde düşük kaldı. Ev tahliyelerini tahmin ederken, en iyi mo-

del 0,22'lik bir korelasyona (YU = %57) ulaştı. Çocuğun birinci derecede bakımından sorumlu kişinin işten çıkarılması ya da iş eğitiminde olması ve çocuğun kendi bildirimine dayalı bir "azim" ölçümü (belli bir amaç için dayanıklılık ve tutkusunu birleştiren bir kişilik özelliği) gibi diğer tekil olay sonuçlarında da benzer değerler bulundu. Bunlar için korelasyonlar 0,17 ve 0,24 arasına denk düşüyordu (YU = % 55-58).

Altı hedef sonuçtan ikisi çok daha öngörülebilir olan toplamalardı. Kestirimci korelasyonlar çocuğun not ortalaması ile 0,44 (YU = %65) ve önceki on iki ay boyunca yaşanan maddi zorluğun özel ölçümüyle 0,48 (YU = %66) idi. Bu ölçüm "Hiç aç kaldın mı?" ya da "Telefon hizmetiniz iptal edildi mi?" sorularının da aralarında olduğu on bir soru üstüne kuruldu. Toplama ölçüler yaygın olarak tekil sonuçların ölçülerinden daha kestirimci ve öngörülebilir olarak bilinirler. Zorluğun başlıca çıkarımı büyük bir kestirimci bilgi yığınının insanların hayatındaki tekil olayların öngörülmesinde yeterli olmadığıdır. Hatta toplamaların öngörüsü bile sınırlıdır.

Bu araştırmada gözlemlenen sonuçlar tipiktir ve sosyal bilimcilerin bildirdiği pek çok korelasyon bu aralığa düşer. Sosyal psikolojide yüzyıl içinde yapılmış 25.000 çalışmayı ve 8 milyon deneği içine alan bir araştırma incelemesi "sosyal psikolojik etkilerin tipik olarak 0,21'e denk bir *r* değeri [korelasyon katsayısı] verdiği" sonucuna vardı. Daha önce bahsettiğimiz yetişkin boyu ve ayak ölçüsü arasındaki 0,60'lık korelasyon gibi daha yüksek korelasyonlar fiziksel ölçümler için yaygındır ama sosyal bilimlerde çok nadirdir. Araştırmacılar davranış bilimi ve bilişsel bilimde 708 çalışmanın bir incelemesi bildirilen korelasyonların sadece %3'ünün 0,50 ve daha fazla olduğunu ortaya koydular.

"İstatistiksel açıdan anlamlı" ve hatta "fazlasıyla anlamlı" olarak sunulan bulguları okumaya alışıksanız, bu tür düşük korelasyon katsayıları sizi şaşırtabilir. İstatistiksel terimler mesleğe yabancı

okuyucular için genellikle yanıltıcıdır ve "anlamlı" bunun en kötü örneği olabilir. Bir bulgu "anlamlı" olarak tanımlandığında, tanımladığı etkinin güçlü olacağı sonucuna varmamalıyız. Bu, basitçe bulgunun tek başına şans ürünü olmasının pek mümkün olmadığı anlamına gelir. Yeterince büyük bir örnekte, bir korelasyon aynı anda hem çok "anlamlı" hem de tartışılmaya değmeyecek kadar küçük olabilir.

Meydan okuma çalışmasındaki tekil sonuçların sınırlı öngörülebilirliği anlamak ve tahmin arasındaki fark hakkında rahatsız edici bir mesaj taşır. Kırılgan Aileler Çalışması bir sosyal bilim hazine sandığı olarak görülür ve gördüğümüz gibi, verileri çok geniş bir araştırma yelpazesinde kullanılmıştır. O araştırmayı üreten bilim insanları elbette çalışmalarının kırılgan ailelerin hayatlarıyla ilgili anlayışı ileriye taşıdığını düşündüler. Ne yazık ki bu ilerleme hissi bireysel hayatlardaki bireysel olaylar hakkında detaylı tahminler yapma becerisiyle örtüşmemişti. Kırılgan Aileler meydan okuması üstüne kaleme alınan çok yazarlı raporun giriş özeti katı bir öğüt içeriyordu: "Araştırmacılar hayat güzergâhlarını, tahminlerin hiçbiri çok isabetli olmadan anladıkları fikriyle barışmalılar."

Anlayış ve Öngörü

Bu kötümser sonucun arkasındaki mantık detaylandırılmayı gerektiriyor. Kırılgan Aileler meydan okumasının yazarları anlamayı öngörüyle (ya da birinin yokluğunu diğerinin yokluğuyla) denk tutarken, *anlama* terimini çok belirli bir anlamda kullanırlar. Kelimenin başka anlamları da var: Matematiksel bir kavramı anladığınızı ya da sevginin ne olduğunu anladığınızı söylediğiniz zaman, muhtemelen belirli tahminler yapma becerisinden bahsetmiyorsunuzdur.

Bununla birlikte, sosyal bilim söyleminde ve çoğu gündelik sohbette, bir şeyi anlama iddiası o şeye neyin *neden olduğunu*

anlama iddiasıdır. Kırılgan Aileler Çalışması'ndaki binlerce değişkeni toplayan ve inceleyen sosyologlar gözlemledikleri sonuçların nedenlerini arıyorlardı. Bir hastaya neyin rahatsızlık verdiğini anlayan doktorlar teşhis ettikleri patolojinin gözlemledikleri semptomların nedeni olduğunu iddia etmektedirler. Anlamak nedensel bir zinciri tarif etmektir. Bir tahminde bulunma becerisi böyle bir nedensel zincirin teşhis edilip edilmediğinin bir ölçüsüdür. Ve korelasyon, kestirimci doğruluğun ölçüsü, ne kadar nedensellik açıklayabildiğimizin ölçüsüdür.

Eğer başlangıç düzeyinde istatistiğe maruz kaldıysanız ve sıklıkla tekrarlanan "korelasyon nedenselliği ima etmez" uyarısını hatırlıyorsanız bu son beyan sizi şaşırtabilir. Örneğin, çocuklarda ayakkabı numarası ile matematiksel beceri arasındaki korelasyonu düşünün: Tahmin edileceği gibi, bir değişken diğerine neden olmaz. Korelasyon ayakkabı ölçüsü ve matematik bilgisinin çocuğun yaşıyla birlikte artmasından doğar. Korelasyon gerçektir ve bir öngörüyü destekler: Bir çocuğun ayaklarının büyük olduğunu biliyorsanız, çocuğun küçük ayaklı olduğunu bilme durumunuza göre daha yüksek bir matematik seviyesi öngörmeniz gerekir. Ama bu korelasyondan nedensel bir bağlantı çıkarmamalısınız.

Öte yandan korelasyon nedenselliği ima etmezken, nedenselliğin korelasyonu ima ettiğini de unutmamalıyız. Nedensel bir bağlantının olduğu her yerde bir korelasyon bulmalıyız. Yetişkinler arasında yaş ve ayakkabı numarası arasında korelasyon bulmazsanız ergenlik çağından sonra yaşın ayakların büyümesine neden olmadığı ve ayakkabı ölçüsündeki farklılıklar için başka yere bakmanız gerektiği sonucuna rahatlıkla varabilirsiniz.

Kısaca nedenselliğin olduğu yerde korelasyon vardır. Buradan nedenselliğin olduğu yerde öngörüde bulunabiliyor olabilmemiz gerektiği –ve bu öngörünün isabetliliği anlamını gelen korelasyonun ne kadar nedenselliğin ölçüsü olduğu– sonucu çıkar. Dolayısıyla Princeton araştırmacılarının çıkarımı şudur: Sosyologların

tahliye gibi olayları öngörme derecesi, 0,22'lik bir korelasyonla ölçüldüğü gibi, bu ailelerin hayat güzergâhları konusunda ne kadar çok –ya da ne kadar az– şey anladıklarının bir göstergesidir. Objektif bilgisizlik sadece öngörülerimize değil, anlayışımıza da bir tavan oluşturur.

Bu durumda, çoğu profesyonelin sahalarını anladıklarını kendilerinden emin bir tavırla iddia ederken kastettikleri nedir? Gözlemledikleri fenomenlere neyin neden olduğu konusunda nasıl beyanda bulunabiliyorlar ve bu konuda nasıl öz güvenli öngörülerde bulunabiliyorlar? Kısaca, profesyoneller –ve hepimiz– neden dünyaya dair objektif bilgisizliklerini hafife alır gibi görünüyorlar?

Nedensel Düşünme

Bu bölümün ilk kısmını okurken kendinize ev tahliyelerine ve kırılgan ailelerde yaşanan diğer hayat olaylarına nelerin neden olduğunu sorduysanız, burada çabalarını anlattığımız araştırmacılarla aynı türde bir düşünce şekline girdiniz demektir. *İstatistiksel düşünme* uyguladınız, kırılgan ailelerin nüfusu gibi bütüne ve onları tanımlayan ortalamaların, farklılıkların, korelasyonların vesaire aralarında olduğu istatistiklere kafa yordunuz. Bireysel vakalara odaklanmadınız.

Akıllarımıza daha doğal olarak gelen farklı bir düşünme şekline burada *nedensel düşünme* diyeceğiz. Nedensel düşünme içinde belli olayların, insanların ve nesnelerin birbirini etkilediği hikâyeler yaratır. Nedensel düşünmeyi tecrübe etmek için kendinizi kıt olanaklarla yaşayan çok sayıda ailenin vakalarını takip eden bir sosyal hizmet görevlisi olarak düşünün. Bu ailelerden birinin, Jonesların evlerinden tahliye edildiğini duydunuz. Bu olaya gösterdiğiniz tepki, Joneslar hakkında bildiklerinizle şekillenir. Şu işe bakın ki ailenin geçimini sağlayan Jessica Jones birkaç ay önce işten çıkarıldı. Başka bir iş bulamadı ve o zamandan beri

kirasını tam olarak ödeyemiyor. Kısmi ödemeler yaptı, bina yöneticisine defalarca dil döktü ve hatta sizden duruma müdahale etmenizi rica etti (müdahale ettiniz ama yönetici etkilenmedi). Bu bağlamda, Jones'ların evlerinden çıkarılmaları hüzünlüdür ama şaşırtıcı değildir. Aslında, bir olaylar zincirinin mantıklı sonu gibi görünür. Gidişatı belli bir trajedinin kaçınılmaz sonu.

Bu kaçınılmazlık duygusuna teslim olduğumuz zaman, her şeyin nasıl kolayca farklı olabileceğini, yoldaki her ayrımda kaderin başka bir yöne sapabileceğini gözden kaçırırız. Jessica işini koruyabilirdi. Hızlı bir şekilde yeni bir iş bulabilirdi. Bir akrabası yardımına koşabilirdi. Siz, sosyal hizmet görevlisi olarak daha etkili bir savunma sağlayabilirdiniz. Bina yöneticisi daha anlayışlı olabilir ve aileye birkaç hafta nefes aldırarak Jessica'nın bir iş bulmasını ve kira açığını kapatmasını mümkün kılabilirdi.

Son biliniyorsa, bu alternatif hikâyeler de şaşırtıcı olmaya ana hikâye kadar uzak olur. Sonuç ne olursa olsun, bir kez gerçekleştiğinde, nedensel düşünme tamamen açıklanabilir, hatta öngörülebilir görünmesine neden olur.

Normallik Vadisini Anlamak

Bu gözlemin psikolojik bir açıklaması var. Bazı olaylar şaşırtıcıdır; ölümcül bir salgın, İkiz Kuleler'e bir saldırı, sonradan bir saadet zinciri olduğu anlaşılan yıldız bir koruma fonu gibi. Özel hayatlarımızda da zaman zaman şoklar yaşarız: bir yabancıya âşık olmak, kardeşimizin genç yaşta ölümü, beklenmedik bir miras gibi. Bir ikinci sınıf öğrencisinin okuldan eve belli bir saatte dönmesi gibi, diğer olaylar aktif olarak beklenen şeylerdir.

Ancak çoğu insan tecrübesi bu iki aşırı ucun arasına denk gelir. Bazen belli bir olayı aktif olarak bekliyoruzdur, bazen de şaşırırız. Ancak çoğu şey, olayların ne tamamen beklendiği ne de özellikle şaşırtıcı olduğu geniş normallik vadisinde gerçekleşir. Örneğin, bu

noktada bir sonraki paragrafta neyin geliyor olabileceği konusunda belirli bir beklentiniz yok. Bir anda Türkçeye geçsek şaşırırsınız ama sizi şoke etmeden söyleyebileceğimiz bir sürü şey var.

Normallik vadisinde olaylar Jonesların tahliyesi gibi gelişir: Beklenmedik olsalar da ve onları öngörememiş olsak da geriye dönüp bakıldığında normal görünürler. Bunun nedeni gerçeği anlama sürecinin geriye dönük olmasıdır. Aktif olarak beklenmeyen bir olay (Jones ailesinin evlerinden çıkarılması) bir aday sebep için (zorlu iş piyasası, katı bina yöneticisi) hafızanın yoklanmasını tetikler. İyi bir anlatı bulunduğu zaman, arama durur. Tam tersi bir sonuçta, araştırma eşit derecede sağlam başka nedenler üretirdi (Jessica Jones'un azmi, bina yöneticisinin anlayışı gibi).

Bu örneklerin resmettiği gibi, normal bir hikâyedeki pek çok olay kendi kendini açıklar. Tahliye hikâyesinin iki farklı versiyonundaki bina yöneticisinin aslında aynı insan olmadığını fark etmiş olabilirsiniz: İlki anlayışsız, ikincisi iyi kalpliydi. Ama yöneticinin karakteriyle ilgili tek ipucunuz karakterini sergilediği davranıştı. Hakkında bildiklerimiz göz önüne alındığında, davranışı tutarlı görünüyor. Olayın nedenini size oluş şekli anlatıyor.

Beklenmedik ama şaşırtıcı olmayan bir olayı bu şekilde açıkladığınızda, sonunda varılan nokta her zaman akla yatkın gelir. Bir hikâyeyi *anlamaktan* kastımız budur ve geri dönüp baktığımızda gerçeği öngörülebilir yapan da budur. Çünkü olay gerçekleşirken kendini açıklar, beklenebileceği yanılsamasına kapılırız.

Daha geniş anlamda olağanüstü, dünyayı anlama hissimiz gözlemlediğimiz olayları açıklayan anlatımlar inşa etme becerimize bağlıdır. Neden arayışı neredeyse her zaman başarıyla sonuçlanır çünkü nedenler dünya konulu sınırsız bir olgular ve inanışlar rezervinden çekilebilir. Örneğin akşam haberlerini dinleyen herkesin bileceği gibi, birkaç büyük borsa hareketi açıklanamamıştır. Aynı haber akışı endekslerde ya bir düşüşü (gergin yatırımcılar

haber karşısında endişelenir) ya da bir yükselmeyi (şen şakrak yatırımcılar iyimserliklerini korur) "açıklayabilir".

Bariz bir neden arayışı başarısız olduğu zaman, başvurduğumuz ilk çare dünya modelimizde bir boşluğu doldurarak bir açıklama üretmek olur. Daha önce bilmediğimiz bir olguya (örneğin, yöneticinin alışılmadık derecede iyi kalpli bir insan olduğuna) bu şekilde ulaşırız. Ancak dünya modelimiz sonucu üretecek şekilde bükülemediğinde, bu sonucu şaşırtıcı olarak yaftalar ve daha detaylı bir açıklama aramaya koyuluruz. Hakiki şaşkınlık sadece rutin geriye dönüp bakma eylemi başarısızlıkla sonuçlandığında yaşanır.

Gerçekliğin bu sürekli nedensel yorumu dünyayı "anlama" şeklimizdir. Hayatı yaşanırken anlama hissimiz normallik vadisinde sürekli bir geriye bakış akışından oluşur. Bu his esasen nedenseldir: Yeni olaylar, bir kez bilinirlik kazanınca, alternatifleri bertaraf eder ve anlatım belirsizliğe pek yer bırakmaz. Geriye bakış üzerine yapılan klasik araştırmalardan bildiğimiz gibi, öznel belirsizlik bir süre var olsa bile, belirsizlik giderildiği zaman, anıları büyük ölçüde silinir.

İçerisi ve Dışarısı

Olaylar hakkında iki düşünme şeklini, istatistik ve nedensel düşünmeyi karşılaştırdık. Nedensel düşünme olayları gerçek zamanda normal ve anormal olarak sınıflandırarak bizi çaba gerektiren düşünmeden korur. Anormal olaylar hem çevrede hem hafızada ilgili bilgi arayışıyla maliyetli çabayı derhâl harekete geçirir. Aktif beklenti –bir şeyin olmasını dikkatle beklemek– de çaba gerektirir. Aksine, normallik vadisindeki olay akışı çok az zihinsel çaba ister. Yolunuzun kesiştiği komşunuz gülümseyebilir ya da düşünceli görünüp sadece başıyla selam verebilir. Geçmişte kararınca sık yaşandılarsa bu iki olay da çok fazla dikkat çekmez. Gülümseme her zamankinden büyük ya da baş selamı her zamankinden daha

yarım yamalaksa kendinizi hafızanızda olası bir neden ararken bulabilirsiniz. Nedensel düşünme bir yandan anormal olayları saptamak için gerekli olan ihtiyatı korurken, gereksiz çaba harcamaktan da kaçınır.

Tam tersine, istatistiksel düşünme çaba ister. Sadece Sistem 2'nin, yavaş ve kasti düşünceyle bağdaştırılan düşünme şeklinin başvurabileceği dikkat kaynaklarını gerektirir. Başlangıç seviyesinin ötesinde, istatistiksel düşünme ayrıca özel eğitim gerektirir. Bu tür düşünme topluluklarla başlar ve bireysel vakaları daha geniş kategorilerin örnekleri olarak kabul eder. Jonesların evlerinden çıkarılmaları, belirli bir olaylar zincirinin sonucu olarak değil, daha önce gözlemlenen Joneslarla aynı kestirimci özellikleri paylaşan vakalara dayanarak, istatistiksel olarak olası (ya da olasılıksız) olarak görülür.

Bu iki görüş arasındaki ayrım bu kitabın kendini tekrar tekrar gösteren temasını oluşturuyor. Tek bir vakada nedensel düşünmeye bel bağlamak, bir öngörülebilir hatalar kaynağıdır. *Dış bakış* olarak da adlandıracağımız istatistiksel bakış açısını benimsemek, bu hatalardan kaçınmanın yoludur.

Bu noktada asıl vurgulamamız gereken, nedensel modun bizlere çok daha doğal gelmesidir. İstatistiksel olarak ele alınması gereken açıklamalar bile kolayca nedensel anlatıma dönüştürülebilir. "Tecrübeleri olmadığı için başarısız oldular", ya da "muhteşem bir liderleri olduğu için başardılar", gibi iddiaları düşünün. Tecrübesiz takımların başarılı olduğu ve muhteşem liderlerin başaramadığı aksi örnekler düşünmek sizin için kolay olurdu. Tecrübe ve muhteşemliğin başarıyla korelasyonları en iyi ihtimalle orta düzeyde, muhtemelen düşüktür. Yine de nedensel bir atıf kolayca yapılır. Nedenselliğin makul olduğu durumlarda zihinlerimiz bir korelasyonu, ne kadar düşük olursa olsun, kolayca nedensel ve açıklayıcı bir güce dönüştürebilir. Muhteşem liderlik başarı için

tatmin edici bir başarı açıklaması olarak, tecrübesizlik de başarısızlık açıklaması olarak kabul edilir.

Alternatifi dünyamızı anlamaktan vazgeçmek olduğunda kusurlu açıklamalara bel bağlamak belki kaçınılmazdır. Ancak nedensel düşünme ve geçmişi anlama yanılsaması geleceğe dair aşırı öz güvenli öngörülere katkı sağlar. Göreceğimiz gibi, nedensel düşünmenin tercih edilmesi ayrıca gürültünün bir hata kaynağı olarak ihmal edilmesine de yol açar çünkü gürültü esasen istatistiksel bir kavramdır.

Nedensel düşünme, sandığımızdan çok daha az öngörülebilir olan bir dünyaya anlam vermemize yardımcı olur. Ayrıca dünyayı neden aslında olduğundan daha öngörülebilir gördüğümüzü de açıklar. Normallik vadisinde, sürprizler ya da tutarsızlıklar yoktur. Gelecek geçmiş kadar öngörülebilir görünür. Ve gürültü ne duyulur ne de görülür.

Anlamanın Sınırlarından Bahsederken

"0,20 civarı korelasyonlar (YU = %56) insani ilişkilerde hayli yaygındır."

"Korelasyon nedenselliği düşündürmez ama nedensellik korelasyonu düşündürür."

"Çoğu normal olay ne beklenir ne de şaşırtıcıdır ve açıklama gerektirmez."

"Normallik vadisinde, olaylar ne beklenir ne de şaşırtıcıdır; sadece kendi kendilerini izah ederler."

"Burada neler olduğunu anladığımızı düşünürüz ama olanı öngörebilir miydik?"

KISIM 4

Gürültü Nasıl Olur?

Gürültünün kökeni nedir? Ya ön yargının? Kararlarımızdaki değişkenliğe ve onları etkileyen ortak hatalara hangi zihinsel mekanizmalar yol açar? Kısaca, gürültünün psikolojisi hakkında ne biliyoruz? Şimdi bu sorulara yöneleceğiz.

Öncelikle hızlı, *Sistem 1* düşünme operasyonlarının neden pek çok yargı hatasından sorumlu olduğunu tanımlayacağız. Bölüm 13'te Sistem 1'in yaygın olarak dayandığı üç önemli yargısal kısayolu sunacağız. Bu kısayolların nasıl gürültü kadar, öngörülebilir yön hatalarına da (istatistiksel ön yargı) neden olduğunu göstereceğiz.

Bölüm 14 eşleştirmeye –Sistem 1'in özel bir operasyonuna– odaklanıyor ve üretebileceği hataları tartışıyor.

Bölüm 15'te bütün yargıların olmazsa olmaz aksesuarına döneceğiz: kararların dayandığı ölçek. Uygun bir ölçek seçiminin iyi kararın ön koşulu olduğunu ve kötü tanımlanmış ve yetersiz ölçeklerin önemli bir gürültü kaynağı olduğunu göstereceğiz.

Bölüm 16, belki de gürültü türlerinin en ilgi çekicisi olan psikolojik kaynağı ele alacak: farklı insanların farklı vakalara tepki örüntüsü. Bireysel kişilikler gibi, bu örüntüler de rastgele

değildir ve çoğunlukla zamanla sabitleşirler ama etkileri kolayca öngörülemez.

Son olarak Bölüm 17'de gürültü ve ögeleri hakkında öğrendiklerimizi özetleyeceğiz. Bu keşif bizi daha önce bahsettiğimiz bir bulmacaya bir cevap sunmaya yönlendirecek: Gürültü her yerde ve her zaman var olmasına rağmen, neden nadiren önemli bir sorun olarak görülür?

BÖLÜM 13

Kısayollar, Ön Yargılar ve Gürültü

Bu kitap sezgisel insan yargısı, sözde kısayollar ve ön yargı programları üstüne yarım yüzyıllık bir araştırmayı kapsıyor. Bu araştırma programının ilk kırk yılı sezgisel düşünmenin hem muhteşem yanlarını hem kusurlarını açıklayan psikolojik mekanizmaları inceleyen *Hızlı ve Yavaş Düşünme* kitabında özetlenmişti. Programın ana fikri zor bir soru sorulan insanların *kısayol* denen basitleştirici işlemler kullanmasıydı. Genel olarak, *Sistem 1 düşünme* olarak bilinen hızlı ve sezgisel düşünmeyle üretilen kısayollar hayli faydalılar ve yeterli cevap üretirler. Ancak bazen sistematik olarak tarif ettiğimiz ön yargılara, öngörülebilir yargı hatalarına yol açarlar.

Kısayol ve ön yargı programı insanların ortak noktalarına değil, farklılıklarına odaklandı. Yargı hatalarına yol açan süreçlerin büyük ölçüde ortak olduğunu gösterdi. Kısmen bu geçmiş yüzünden, psikolojik ön yargı kavramına aşina olan insanlar genellikle, her zaman *istatistiksel ön yargı* (bu kitapta çoğunlukla gerçekten aynı yönde sapan ölçüm ya da yargılar anlamında kullandığımız bir terim) da ürettiğini var sayarlar. Gerçekten de psikolojik ön yargılar yaygın olarak paylaşıldıklarında istatistiksel ön yargı

üretirler. Ancak psikolojik ön yargılar, yargıçlar farklı şekillerde ya da farklı derecelerde ön yargılı olduklarında sistem gürültüsü yaratırlar. Elbette, istatistiksel ön yargıya ya da gürültüye neden olsunlar ya da olmasınlar, psikolojik yargılar her zaman hata yaratır.

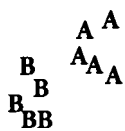
Ön Yargıları Teşhis Etmek

Karar ön yargıları sıklıkla gerçek bir değere referansla teşhis edilir. Hatalar çoğunlukla biri yerine bir diğer yöndeyse öngörücü yargılarda ön yargı var demektir. Örneğin insanlar bir projeyi tamamlamalarının ne kadar süreceği konusunda tahminde bulunurken, tahminlerinin ortalaması genellikle aslında ihtiyaç duyacakları süreden çok daha düşük olur. Bu tanıdık psikolojik ön yargı *planlama yanılgısı* olarak bilinir.

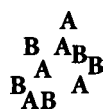
Gerçi, genellikle yargıların karşılaştırılabileceği gerçek bir değer yoktur. İstatistiksel ön yargının sadece gerçek değer bilindiğinde saptanabildiğini defalarca vurgulamamızdan ötürü, gerçek bilinmez olduğunda psikolojik ön yargıların nasıl incelenebileceğini merak ediyor olabilirsiniz. Cevap şudur: Araştırmacılar psikolojik bir ön yargıyı, yargıyı etkilememesi gereken bir faktörün yargının üstünde istatistiksel bir etkisi olduğunu ya da yargıyı etkilemesi gereken bir faktörün yargı üstünde istatistiksel etkisi olmadığını gözlemleyerek teyit ederler.

Bu yöntemi görselleştirmek için, yeniden atış poligonu benzetmesine dönelim. A ve B takımlarının atışlarını yaptığını ve hedef tahtasının arkasına baktığımızı varsayalım (Şekil 12). Bu örnekte, boğa gözünün nerede olduğunu bilmiyorsunuz (gerçek değer bilinmiyor). Bu nedenle iki takımın hedefin merkezine göre ne kadar ön yargılı olduklarını bilmiyorsunuz. Öte yandan Panel 1'de size iki takımın aynı boğa gözüne nişan aldığını ve Panel 2'de A Takımı'nın bir boğa gözünü, B Takımı'nın ise başka bir boğa gözünü hedef aldığı söylendi.

Bir hedefin yokluğuna rağmen, her iki panel sistematik ön yargı kanıtı sağlar. Panel 1'de iki takımın atışları aynı olmaları gerekirken farklılık gösteriyor. Bu örüntü iki yatırımcı grubunun büyük ölçüde aynı olan ama farklı bir fontla, farklı kâğıtlara basılmış iş planlarını okuduğu bir deneyde göreceğiniz örüntüye benzer. Bu alakasız detaylar yatırımcıların yargısında bir fark yaratıyorsa psikolojik ön yargı var demektir. Şık font ve parlak kâğıttan etkilenen yatırımcılar çok fazla pozitif mi yoksa daha kaba versiyonu okuyanlar çok negatif mi bilmeyiz. Ama öyle olmaması gerekirken, yargılarının farklı olduğunu biliriz.



Panel 1
Aynı boğa gözüne nişan alıp farklı noktalara isabet ettirilmiş



Panel 2
Farklı boğa gözlerine nişan alıp aynı bölgeye isabet ettirilmiş

ŞEKİL 12: *Hedef tahtasının arkasına bakmak, ön yargı testi için deney sağlıyor*

Panel 2 tam zıt bir fenomeni sergiliyor. Takımlar farklı açılardan nişan aldıkları için, atış öbekleri ayrı olmalı ama aynı noktada toplanmışlar. Örneğin, iki insan grubuna size Bölüm 4'te Michael Gambardi hakkında sorulan sorunun üstünde biraz oynanarak sorulduğunu düşünün. Bir gruptan, sizde olduğu gibi, Gambardi'nin iki yıl sonra aynı işte olma ihtimalini tahmin etmeleri istensin. Diğerinden ise üç yıl sonra aynı işte olma olasılığını tahmin etmeleri istensin. İki grup farklı sonuçlara varmalıdır, çünkü tahmin edileceği gibi, işinizi üç yıl sonra kaybetme olasılığı iki yıl içinde kaybetmekten daha fazladır. Ancak kanıtlar iki grubun olasılık tahminlerinin çok az farklılık sergileyeceğini

gösteriyor, tabii sergilerlerse. Cevaplar açık bir şekilde farklı olmalı ama değiller ve bu, yargıları etkilemesi gereken bir faktörün görmezden gelindiğini düşündürüyor. (Bu psikolojik ön yargıya *kapsam duyarsızlığı* denir.)

Sistematik yargı hataları pek çok alanda sergilenmiştir ve *ön yargı* terimi artık iş, politika, politika oluşturma ve hukuk dâhil pek çok alanda kullanılmaktadır. Sözcük yaygın olarak kullanıldığı için anlamı geniştir. Burada kullandığımız bilişsel tanıma (psikolojik bir mekanizmaya ve bu mekanizmanın tipik olarak ürettiği hataya atfen) ek olarak sözcük bir kişinin belli bir gruba karşı ön yargılı olduğunu (örneğin cinsiyetçi ön yargılar ya da ırksal ön yargıları) öne sürmek için de kullanılır. Ayrıca birinin bir çıkar çatışması ya da politik bir görüş nedeniyle ön yargılı olduğunu okuduğumuzda olduğu gibi kişinin özel bir sonucu kayırdığı anlamına da gelebilir. Yargı hatalarının psikolojisini ele alırken bu ön yargı türlerini de işin içine katıyoruz çünkü bütün psikolojik ön yargılar hem istatistiksel ön yargıya hem de gürültüye neden olur.

Öte yandan şiddetle karşı durduğumuz bir kullanım var. Bu kullanımda maliyetli başarısızlıklar belirtilmemiş "ön yargılara" bağlanır ve hatanın kabullenilmesine "karar alma sürecimizdeki ön yargıları bertaraf etmek için çok çalışma" vaatleri eşlik eder. Bu beyanlar "hatalar yapıldı" ve "daha iyisini yapmak için çok çalışacağız"ın ötesinde bir anlam ifade etmez. Evet, bazı başarısızlıklara, gerçekten, belirli psikolojik ön yargılarla bağdaştırılan öngörülebilir hatalar neden olur ve müdahalelerin yargılar ve kararlarda ön yargıyı (ve gürültüyü) azaltmasının mümkün olduğuna inanıyoruz. Ancak istenmeyen her sonucun yükünü ön yargılara yüklemek değersiz bir açıklamadır. *Ön yargı* kelimesinin belirli ve teşhis edilebilir hatalara ve onları üreten mekanizmalara ayrılmasını tavsiye ediyoruz.

İkame

Bu kısayol sürecini tecrübe etmek için, kısayollar ve ön yargılar yaklaşımının çeşitli temel temalarını temsil eden aşağıdaki soruları yanıtlayarak kendinizi deneyin. Her zamanki gibi, kendi cevaplarınızı üretirseniz örnekten daha fazla fayda sağlarsınız.

Bill otuz üç yaşında. Zeki ama yaratıcı değil, zorlantılı ve genel olarak cansız. Okulda matematikte güçlü ama sosyal ve insani bilimlerde zayıftı.

Aşağıda Bill'in mevcut durumu için sekiz olasılık listesini bulacaksınız.

Lütfen listeyi gözden geçirin ve en yüksek olasılığı verdiğiniz iki tanesini seçin.

- ❑ *Bill hobi olarak poker oynayan bir doktor.*
- ❑ *Bill mimar.*
- ❑ *Bill muhasebeci.*
- ❑ *Bill hobi olarak cazla uğraşıyor.*
- ❑ *Bill hobi olarak sörf yapıyor.*
- ❑ *Bill muhabir.*
- ❑ *Bill hobi olarak cazla uğraşan bir muhasebeci.*
- ❑ *Bill hobi olarak dağcılıkla uğraşıyor.*

Şimdi listenin üstünden geçin ve Bill'in o kategorideki tipik bir insana en çok benzediği iki kategoriyi seçin. Daha önce seçtiklerinizle aynı ya da farklı kategorileri seçebilirsiniz.

Olasılık ve benzerlik açısından en yüksek olarak aynı kategorileri seçiğinizden neredeyse eminiz. Bu güvenimizin nedeni çok sayıda deneyin insanların iki soruya aynı cevapları verdiğini göstermiş olması. Ancak benzerlik ve olasılık aslında çok farklı şeylerdir. Örneğin kendinize aşağıdaki beyanlardan hangisinin daha mantıklı olduğunu sorun.

- ❑ *Bill hobi olarak cazla uğraşan insan görüşüme uyuyor.*
- ❑ *Bill hobi olarak cazla uğraşan muhasebeci görüşüme uyuyor.*

Bu beyanların ikisi de uygun değil ama biri diğerinden açık şekilde daha az korkunç. Bill'in hobi olarak cazla uğraşan bir muhasebeciyle, hobi olarak cazla uğraşan bir insanla olduğundan daha fazla ortak noktası var. Şimdi şunu düşünün: Aşağıdakilerden hangisi daha yüksek olasılıklı?

- ❑ *Bill hobi olarak cazla uğraşıyor.*
- ❑ *Bill hobi olarak cazla uğraşan bir muhasebeci.*

İkinci cevabı seçmek daha cazip görünebilir ama mantığınız buna izin vermez. Bill'in hobi olarak cazla uğraşma olasılığı onun cazla uğraşan bir muhasebeci olmasından daha yüksek *olmalıdır.* Venn şemalarınızı hatırlayın! Bill cazla uğraşıyorsa ve muhasebeciyse, kesinlikle cazla uğraşıyordur. Bir tanıma detay katmak, onu daha temsilî ve dolayısıyla daha "uygun" kılmasına rağmen, bu örnekte olduğu gibi, sadece olasılığını azaltır.

Yargı kısayolları teorisi insanların bazen daha zor bir soruya cevap verirken, daha kolay bir sorunun cevabını kullanabileceklerini öne sürer. Yani, hangi soru daha kolay cevaplanır: "Bill tipik bir amatör caz sanatçısına ne kadar benziyor?" mu, "Bill'in amatör bir caz sanatçısı olması ne kadar olası?" mı? Güçlü bir onayla, benzerlik sorusu daha kolaydır ve bu, insanların olasılığı değerlendirmeleri istendiğinde, o sorunun cevabını verme ihtimalini artırır.

Artık kısayollar ve ön yargılar programının temel fikrini tecrübe ettiniz: Zor bir soruyu cevaplamanın kısayolu daha kolay bir soruya cevap bulmaktır. Bir sorunun diğerinin yerini alması, psikolojik ön yargılar denilen öngörülebilir hatalara neden olur.

Bu ön yargı türü Bill örneğinde görülür. Bir benzerlik yargısı bir olasılık yargısının yerini aldığında, hataların olması kaçınılmazdır çünkü olasılık özel bir mantıkla sınırlıdır. Özellikle, Venn şemaları benzerlikte değil, olasılıkta geçerlidir. Birçok insanın yaptığı öngörülebilir mantık hataları buradan kaynaklanır.

Göz ardı edilen istatistiksel özelliğe bir başka örnek olarak, Bölüm 4'te Gambardi sorusunu nasıl düşündüğünüzü hatırlayın. Çoğu insan gibiyseniz, Michael Gambardi'nin başarı şansıyla ilgili değerlendirmeniz tamamen size onun hakkında söylediklerine dayanmıştır. Daha sona bu tarifi başarılı bir CEO imajıyla eşleştirmeye çalışmışsınızdır.

Rastgele seçilmiş bir CEO'nun iki yıl sonra aynı pozisyonu koruması olasılığını düşünmek hiç aklınıza geldi mi? Muhtemelen hayır. Bu *baz oran bilgisi*ni bir CEO olarak ayakta kalmanın zorluğunun bir ölçüsü olarak düşünebilirsiniz. Bu yaklaşım tuhaf geldiyse belli bir öğrencinin bir sınavı geçme olasılığını nasıl tahmin edeceğinizi düşünün. Elbette sınavda başarısız olan öğrencilerin oranı önemlidir ve size sınavın zorluğu hakkında bir gösterge sağlar. Aynı şekilde, CEO'ların yerlerini koruma baz oranı da Gambardi sorusu açısından önemlidir. Her iki soru da dış bakış adını verdiğimiz şeyi takınmanın örnekleridir. Bu bakış açısından baktığınızda, öğrenciyi ya da Gambardi'yi benzer örnekler sınıfının bir üyesi olarak görürsünüz. Odak vaka üzerinde nedensel olarak düşünmek yerine, sınıfı istatistiksel açıdan düşünürsünüz.

Dış bakışı benimsemek çok büyük bir fark yaratabilir ve önemli hataları önleyebilir. Birkaç dakikalık bir araştırma Amerikan şirketlerindeki CEO değişiminin yıllık yaklaşık %15 civarında olduğunu ortaya koyuyor. Bu istatistik göreve gelen ortalama bir CEO'nun iki yılın sonunda o görevde olma olasılığının kabaca %72 olduğunu düşündürüyor. Elbette, bu oran sadece bir başlangıç noktası ve Gambardi örneğindeki detaylar nihai tahmininizi etkileyecektir. Ama sadece Gambardi hakkında söylenenlere odaklandıysanız,

kilit önem taşıyan bir bilgiyi ihmal etmişsinizdir demektir. (Tam açıklama: Gambardi örneğini gürültülü yargıyı resmetmek için yazdık: Burada tarif etiğimiz şeyin aynı zamanda, *baz oran ihmali* adı verilen önemli bir ön yargı örneği de olduğunu haftalar sonra fark ettik. Baz oranlarını düşünmek bu kitabın yazarları için diğer insanlar için olduğundan daha otomatik değil.)

Bir sorunun yerine başka bir soru koymak sadece benzerlik ve olasılıkla kısıtlı değil. Bir diğer örnek, sıklık yargısının yerine örneklerin akla gelme kolaylığı izleniminin konmasıdır. Örneğin, uçak kazası ve kasırga risklerinin algısı, bu tür olayların kamuya mal olan örneklerinden kısa süre sonra yükselir. Teoride bir risk yargısı uzun vadeli ortalamaya dayanmalıdır. Gerçekte, yakın hadiselere daha fazla ağırlık verilir çünkü kolayca akla gelirler. Örneklerin akla ne kadar kolay geldiği yargısını sıklık değerlendirmesinin yerine koymak *elverişlilik kısayolu* olarak bilinir.

Zor bir yargının yerine basit bir yargıyı koymak bu örneklerle sınırlı değildir. Aslında çok yaygındır. Daha kolay bir soruyu cevaplamak, sizi şaşırtabilecek bir soruyu cevaplamak için genel amaçlı bir prosedür olarak düşünülebilir. Aşağıdaki soruların her birini nasıl daha kolay bir yedeğini kullanarak cevaplama eğiliminde olduğumuzu düşünün.

İklim değişikliğine inanıyor muyum?

Var olduğunu söyleyen insanlara güveniyor muyum?

Bu cerrahın becerikli olduğunu düşünüyor muyum?

Bu birey öz güven ve yetkiyle konuşuyor mu?

Proje zamanında tamamlanacak mı?

Şu anda plana göre ilerliyor mu?

Nükleer enerji gerekli mi?

Nükleer kelimesi beni irkiltiyor mu?

Hayatımdan bütün hâlinde hoşnut muyum?

Şu anda ruh hâlim ne?

Sorudan bağımsız olarak bir sorunun yerine bir başkasını koymak, kanıtın farklı yönlerine uygun ağırlıklarını vermeyen bir cevaba yol açacaktır ve kanıtın hatalı tartılması kaçınılmaz olarak hataya neden olur. Örneğin, hayat memnuniyetiyle ilgili soruya eksiksiz bir cevap, açıkça mevcut ruh hâlinizden daha fazlasını hesaba katmayı gerektirir ama kanıtlar ruh hâline gereğinden fazla ağırlık verildiğini düşündürüyor.

Aynı şekilde, olasılığın yerine benzerliği koymak, benzerlik yargısında oldukça alakasız olan baz oranların göz ardı edilmesine yol açar. Ve bir şirketin değerini değerlendirirken bir iş planını sunan belgenin estetiğindeki alakasız farklılıklar gibi faktörlere çok az değer verilmeli, hatta hiç değer verilmemelidir. Bu faktörlerin yargı üstünde yaratacakları her tür etki kanıtın yanlış tartılmasına yansıyacak ve hata üretecektir.

Sonuç Ön Yargıları

Star Wars filmlerinin üçüncüsü *Jedi'nin Dönüşü*'nün senaryosunun geliştirilmesinin kilit bir anında, serinin arkasındaki deha George Lucas büyük iş birlikçisi Lawrence Kasdan'la hararetli bir tartışma yaşadı. Kasdan Lucas'a "Luke'u öldürmesini ve Leia'yı başa geçirmesini" şiddetle öneriyordu. Lucas bu fikri şiddetle reddetti. Kasdan Luke yaşayacaksa bir başka önemli karakterin ölmesi gerektiğini söyledi. Lucas bu fikre de karşı çıktı, "Sürekli birilerini öldürerek ilerleyemezsin." diye ekledi. Kasdan sinemanın doğasıyla ilgili içten bir iddiayla karşılık verdi. Lucas'a, "yol üstünde sevilen kaybedilirse filmin duygusal ağırlığının arttığını, yolculuğun etkisinin büyüdüğünü" anlattı.

Lucas'ın tepkisi hızlı ve dolambaçsız oldu: "Bu hoşuma gitmiyor ve buna inanmıyorum."

Buradaki düşünce süreci cazla uğraşan muhasebeci Bill'i düşünürken tecrübe ettiğiniz süreçten hayli farklı görünüyor. Lucas'ın

cevabını tekrar okuyun: "Hoşlanmamak", "inanmamak"tan önce geliyor. Lucas Kasdan'ın önerisine otomatik bir tepki verdi. O tepki yargısını motive etmeye yardımcı oldu (sonunda haklı çıksa bile).

Bu örnek, *sonuç ön yargısı* ya da *peşin hüküm* dediğimiz farklı bir ön yargı türünü temsil ediyor. Lucas gibi, bizler de yargı sürecine genellikle belli bir sonuca varma eğilimiyle başlarız. Bunu yaptığımız zaman, hızlı ve sezgisel düşünmemizin yani Sistem 1'in bir sonuç önermesine izin veririz. Ya hızla o sonuca varırız ya da bilgi toplama ve birleştirme sürecini atlarız ya da ön yargımızı destekleyecek argümanlar bulmak için Sistem 2'yi (yavaş ve hesaplı düşünmeyi) harekete geçiririz. Bu durumda, kanıtlar seçici ve çarpıtılmış olur: *Onay ön yargısı* ve *arzu edilebilirlik ön yargısı* nedeniyle, sırasıyla, zaten inandığımız ya da doğru olmasını dilediğimiz bir yargıyı kayırmak için seçici şekilde kanıt toplama ya da yorumlama eğilimi sergileriz.

İnsanlar genellikle yargılarına akla uygun gerekçeler uydurur ve inançlarının sebebinin onlar olduğunu gerçekten düşünürler. Ön yargının rolüyle ilgili iyi bir test, görüntüde inancımızı destekleyen argümanların geçersizliğinin bir anda ispatlandığını hayal etmektir. Örneğin, Kasdan Lucas'a "sürekli birilerini öldürerek ilerleyemezsin"in sağlam bir argüman olmadığını belirtebilirdi. *Romeo ve Juliet*'in yazarı Lucas'a katılmazdı ve *The Sopranos*'un ya da *Game of Thrones*'un yazarları ölümlerin aleyhine karar alsalardı iki dizi de muhtemelen ilk sezonlarında iptal edilirdi. Ama güçlü bir karşı argümanın Lucas'ın fikrini değiştirmeyeceğine bahse girebiliriz. Bunun yerine, muhtemelen kendi yargısını destekleyecek başka argümanlar uydururdu. (Örneğin, "Star Wars farklı.")

Baktığımız her yerde ön yargılar aşikârdır. Lucas'ın tepkisi gibi, sık sık duygusal bir ögeleri olur. Psikolog Paul Slovic buna *duygu temelli kısayol* adını vermiştir. İnsanlar ne düşündüklerini duygularına danışarak belirlerler. Yeğlediğimiz politikacılar hakkında çoğu şeyi severiz ve sevmediğimiz politikacıların

görüntüsünden ve sesinden bile hoşlanmayız. Akıllı şirketlerin markalarını olumlu bir hisle birleştirmek için çok çabalamasının nedenlerinden biri de budur. Profesörler sıklıkla hocalıklarının yüksek notla değerlendirildiği yıllarda, öğrencilerin ders materyaline de yüksek not verdiklerini fark ederler. Öğrenciler hocadan çok hoşlanmadıkları yıllarda ise aynı okuma materyallerine de düşük puan verirler. İşin içine duygu katılmadığı zaman da aynı mekanizma iş başındadır: İnancınızın gerçek nedenleri ne olursa olsun, akıl yürütme şekli yanlış olduğunda bile, onu destekleyen her argümanı kabul etmeye eğilimli olursunuz.

Sonuç ön yargısının daha incelikli bir örneği, keyfî bir sayının sayısal bir karar alması gereken insanlar üstündeki etkisi anlamına gelen *çapa atma etkisidir.* Tipik bir örnekte, size yabancısı olduğunuz bir şarap gibi fiyatlarını tahmin etmenin kolay olmadığı birkaç ürün sunulur. Sizden Sosyal Güvenlik numaranızın son iki hanesini yazmanız ve söz konusu şarap için o miktarı ödeyip ödemeyeceğinizi belirtmeniz istenir. Son olarak sizden ödeyeceğiniz en yüksek tutarı belirtmeniz rica edilir. Sonuçlar Sosyal Güvenlik numaranıza çapa atmanın nihai satın alma fiyatınızı etkileyeceğini gösterir. Bir çalışmada Sosyal Güvenlik numarası yüksek bir çapa üreten (seksen doların üstünde) insanlar, çapası düşük olanların (seksen doların altı) üç katı fazla ücret ödemeye razı olduklarını belirttiler.

Şüphesiz, Sosyal Güvenlik numaranız bir şarabın ederiyle ilgili yargınız üstünde bu kadar büyük etki yaratmamalı ama yaratıyor. Çapa atma son derece sağlam bir etkidir ve pazarlıklarda sık sık kasten kullanılır. İster bir pazarda pazarlık edin, ister karmaşık bir ticari alışveriş için masaya oturun, ilk teklifi yapmak, muhtemelen lehinize olacaktır çünkü çapanın alıcısı istemsiz olarak teklifinizin hangi yönlerden akla yatkın olduğunu düşünmeye çekilecektir. İnsanlar her zaman duyduklarına anlam vermeye çalışırlar; mantıksız

bir sayıyla karşılaştıklarında otomatik olarak akıllarına sayının mantıksızlığını azaltacak değerlendirmeler getirirler.

Aşırı Tutarlılık

Üçüncü bir ön yargı türünü tecrübe etmenize yardımcı olacak bir deney daha. Bir üst düzey yönetici pozisyonu için, bir adayın tanımını okuyacaksınız. Tanım, her biri bir karta yazılmış dört sıfattan oluşuyor. Kart destesi karıştırıldı. İlk iki kartta şu iki kelime var:

Zeki. Israrcı.

Yargıyı bilgi tamamlanana kadar askıya almak mantıklı olabilirdi ama öyle olmadı, adaya dair bir değerlendirmeniz oldu bile ve bu değerlendirme olumlu. Bu yargı basitçe gerçekleşti. Süreç üstünde hiçbir kontrolünüz yoktu ve yargının askıda kalması bir seçenek değildi.

Daha sonra, son iki kartı çekiyorsunuz. Şimdi, tanımın tamamına sahipsiniz.

Zeki, Israrcı, Kurnaz, Prensipsiz

Artık değerlendirmeniz adayın lehine değil ama yeterince değişmedi. Karşılaştırmak için kartların farklı bir şekilde karıştırılmasının üreteceği aşağıdaki tanımı düşünün.

Prensipsiz, Kurnaz, Israrcı, Zeki

Bu ikinci tanım aynı sıfatlardan oluşuyor ama takdim edilme sıraları nedeniyle ilkinden açıkça daha az cazip görünüyor. *Kurnaz* sözcüğü *Zeki* ve *Israrcı*'yı takip ettiğinde sadece kısmen olumsuzdu

çünkü yöneticinin iyi niyetli olduğuna hâlâ (ve sebepsiz) inanıyorduk. Ancak *Kurnaz* kelimesi *Prensipsiz*'in arkasından geldiği zaman, korkunç bir şekil alıyor. Bu bağlamda, ısrarcılık ve zekâ olumlu olmaktan çıkıyor; kötü bir insanı daha da tehlikeli kılıyorlar.

Bu deney *aşırı tutarlılığı* temsil ediyor: Tutarlı izlenimleri çok çabuk oluşturur ve onları değiştirmekte yavaş kalırız. Bu örnekte, çok az kanıt ışığında ve derhâl adaya karşı olumlu bir tavır geliştirdik. Onay ön yargısı (bir ön yargımız olduğunda, bizi çelişkili kanıtı tamamen yok saymaya yönlendiren eğilim) sonraki verilere olması gerekenden daha az önem vermemize neden oldu. (Bu fenomeni tanımlayan bir diğer terim de *halo etkisidir.* Çünkü aday, ilk izlenimin olumlu "halosuyla" değerlendirildi. Bölüm 24'te halo etkisinin işe alma kararlarında ciddi bir sorun oluşturduğunu göreceğiz.)

Bir başka örnek daha. Amerika Birleşik Devletleri'nde, kamu görevlileri zincir restoranların tüketicilerin örneğin çizburgerlerle, hamburgerlerle ve salatalarla ilintili kalori miktarlarını görebilmesi için, kalori etiketleri kullanmalarını talep etti. Tüketiciler o etiketleri gördükten sonra seçimlerini değiştirirler mi? Kanıtlar tartışmalı ve karmaşıktır. Ancak aydınlatıcı bir çalışmada, kalori etiketleri yiyecek maddesinin sağ tarafı yerine sol tarafına yerleştirildiğinde, tüketicilerin etkilenme olasılığının arttığı anlaşıldı. Kaloriler solda olduğunda tüketiciler önce bu bilgiyi alıyor ve ürünü görmeden, "ne çok kalori!" ya da "o kadar da kalorili değilmiş!" diye düşünüyorlar. Başlangıçtaki olumlu ya da olumsuz tepki seçimlerini büyük ölçüde etkiliyor. Aksine, anlaşılıyor ki insanlar önce yiyecek maddesini gördükleri zaman, kalori etiketini görmeden, "enfes!" ya da "o kadar da harika değil!" diye düşünüyorlar. Burada da yine, ilk tepkileri tercihlerini büyük ölçüde etkiliyor. Bu hipotez yazarların, İbranice konuşan ve dolayısıyla sağdan sola doğru okuyan tüketicilerde, kalori etiketinin sol yerine sağ tarafa konmasının etkiyi önemli ölçüde artırdığı bulgusuyla desteklendi.

Genel olarak, sonuçlara hızlı bir şekilde varır ve onlara tutunuruz. Görüşlerimizi kanıtlar üstüne kurduğumuzu düşünürüz ama değerlendirdiğimiz kanıt ve onu yorumlama şeklimizin ilk hızlı yargımıza uyacak şekilde, en azından bir dereceye kadar çarpıtılması olasıdır. Sonuç olarak aklımızda su yüzüne çıkan genel hikâyenin tutarlılığını koruruz. Bu süreçte, sonuçlar doğruysa elbette sorun yoktur. Ancak başlangıçtaki değerlendirme hatalıysa, çelişkili kanıtlar karşısında o değerlendirmeye bağlı kalma eğiliminin hataları katlayarak büyütme olasılığı yüksektir. Ve bu etkiyi kontrol etmek güçtür, çünkü duyduğumuz ya da gördüğümüz bilgiyi yok saymak imkânsız ve unutmak genellikle zordur. Mahkemelerde yargıçlar bazen jüri üyelerinden duydukları kabul edilemez bir kanıt parçasını duymazdan gelmelerini ister ama bu gerçekçi bir talimat değildir (gerçi açıkça bu kanıta dayanan argümanların reddedilebileceği jüri müzakeresinde işe yarayabilir).

Psikolojik Ön Yargılar Gürültüye Neden Olur

Farklı şekillerde işleyen üç ön yargı türünü kısaca sunduk: Kanıtların yanlış tartılmasına neden olan ikame ön yargısı, kanıtın atlanmasına ya da çarpıtılmış hâliyle hesaba katılmasına neden olan sonuç ön yargıları ve ilk izlenimlerin etkisini artıran ve çelişkili bilginin etkisini azaltan aşırı tutarlılık. Her üç ön yargı türü de elbette, istatistiksel ön yargı üretebilir. Ayrıca gürültü de üretebilirler.

İkameyle başlayalım. Çoğu insan Bill'in muhasebeci olma olasılığını profilinin muhasebeci stereotipine benzerliğiyle yargılar. Bu deneyde sonuç ortak bir ön yargıdır. Cevap veren herkes aynı hatayı yapıyorsa, ortada gürültü yoktur. Ancak ikame her zaman böyle bir oy birliği üretmez. "İklim değişikliği var mı?" sorusunun yerine, "Gerçek olduğunu söyleyen insanlara güveniyor muyum?" sorusunu koyduğumuz zaman, cevabın bireyin sosyal çevresine,

tercih ettiği bilgi kaynaklarına, politik eğilimine vs. bağlı olarak kişiden kişiye farklılık göstereceğini görmek kolaydır. Aynı psikolojik ön yargı çeşitli yargılar ve kişiler arası gürültü yaratır.

İkame ayrıca bir durum gürültüsü kaynağı da olabilir. Hayat memnuniyetiyle ilgili bir soru kişinin o andaki ruh hâline dayanılarak cevaplanırsa, cevap aynı kişi için kaçınılmaz olarak andan ana değişecektir. Mutlu bir sabahın arkasından sıkıntılı bir öğleden sonra gelebilir ve değişen ruh hâlleri, mülakatçının soruyu ne zaman sorduğuna bağlı olarak zaman içinde çok farklı hayat memnuniyeti bildirimlerine neden olabilir. Bölüm 7'de, ucu psikolojik ön yargılara gidebilen durum gürültüsü örneklerini ele almıştık.

Yargıya varmadan önce verilen hükümler de ön yargı ve gürültü üretebilir. Giriş bölümünde değindiğimiz bir örneğe dönelim: yargıçların onay verdiği sığınma talebi yüzdelerindeki şaşırtıcı uyuşmazlık. Bir yargıç başvuruların %5'ini onaylarken aynı mahkemedeki bir başka yargıç %88'ini kabul ettiğinde, farklı yönlerde ön yargılı olduklarından emin olabiliriz. Daha geniş bir bakış açısıyla ön yargılardaki bireysel farklılıklar çok büyük sistem gürültüsüne neden olabilir. Elbette, sistemin kendisi de çoğu yargıcın ya da bütün yargıçların benzer şekilde ön yargılı olacağı derecede ön yargılı olabilir.

Son olarak aşırı tutarlılık bilgi sekansının ve bilgiye atfedilen anlamın bütün (ya da çoğu) yargıçlar için aynı olup olmamasına bağlı olarak, ön yargı ya da gürültü üretebilir. Örneğin, görüntüsü çoğu işe alım uzmanında erken bir olumlu izlenim yaratan fiziksel olarak çekici bir aday düşünün. Dış görüntü adayın değerlendirildiği pozisyon için önemsizse, bu olumlu "halo" ortak bir hataya –ön yargıya– yol açacaktır.

Öte yandan birçok karmaşık karar özünde gelişigüzel bir sırayla ulaşan bilginin derlenmesini gerektirir. Bölüm 2'deki sigorta eksperlerini düşünün. Bir taleple ilgili verilerin geliş sırası bir

eksperden diğerine ve bir vakadan diğerine değişerek ilk izlenimlerde rastgele çeşitliliğe neden olur. Aşırı tutarlılık bu rastgele çeşitliliğin nihai yargılarda rastgele çarpıtılmalara neden olacağı anlamına gelir. Etkisi sistem gürültüsü olur.

Kısaca, psikolojik ön yargılar bir mekanizma olarak evrenseldir ve sıklıkla ortak hatalar üretirler. Ancak ön yargılarda büyük bireysel farklılıklar (farklı peşin hükümler) olduğunda ya da ön yargıların etkisi bağlama göre değiştiğinde (farklı tetikleyiciler), gürültü olur.

Ön yargı da gürültü de hata yaratır; bu, psikolojik ön yargıları azaltan her şeyin yargıyı iyileştireceği anlamına gelir. Ön yargıların giderilmesi konusuna Kısım 5'te döneceğiz. Ancak şimdilik, yargı sürecini keşfe devam edeceğiz.

Kısayollar, Ön Yargılar ve Gürültüden Bahsederken

"Psikolojik ön yargılarımız olduğunu biliyoruz ama her hatanın suçunu belirsiz 'ön yargılara' atma arzusuna direnmeliyiz."

"Asıl cevaplamamız gereken sorunun yerine daha kolayını koyduğumuzda, hataların yapılması kaçınılmazdır. Örneğin olasılığı benzerlik gibi değerlendirdiğimizde baz oranı yok sayarız."

"Peşin hükümler ve diğer sonuç ön yargıları insanları kanıtları başlangıçtaki pozisyonlarının lehine çarpıtmaya iter."

"Hızlı bir şekilde izlenim edinir ve çelişen bilgi geldiğinde bile, o izlenimlere sıkı sıkı tutunuruz. Bu eğilime aşırı tutarlılık denir."

"Çok fazla insan aynı ön yargıyı paylaşırsa psikolojik ön yargılar, istatistiksel ön yargıya neden olur. Ancak pek çok örnekte insanların ön yargıları farklılık sergiler. Bu örneklerde psikolojik ön yargılar sistem gürültüsü yaratır."

BÖLÜM 14

Eşleştirme Operasyonu

Gökyüzüne bakın. İki saat içinde yağmur yağma olasılığı nedir? Muhtemelen bu soruyu cevaplamakta sıkıntı yaşamadınız. Vardığınız yargı, örneğin birazdan yağmur yağma olasılığının "çok yüksek" olduğu yargısı çaba gerektirmeden üretildi. Gökyüzünün karanlığıyla ilgili değerlendirmeniz bir şekilde, bir olasılık yargısına dönüştürüldü.

Az önce yaptığınız şey *eşleştirmenin* giriş düzeyinde bir örneğiydi. Yargıyı öznel bir izlenime (ya da bir izlenimin bir yönüne) bir ölçek üzerinde değer atayan bir işlem olarak tanımladık. Eşleştirme o işlemin temel bir parçasıdır. "1 ile 10 arasında, ruh hâlin ne kadar iyi?" ya da "Lütfen bu sabahki alışveriş tecrübenize bir ile beş arasında bir yıldız verin" sorularını yanıtlarken yaptığınız şey eşleştirmedir: Göreviniz yargı ölçeğinde ruh hâlinize ya da tecrübenize uyan bir değer bulmaktır.

Eşleştirme ve Tutarlılık

Bir önceki bölümde Bill'le tanıştınız ve işte yine karşınızda. "Bill otuz üç yaşında. Zeki ama yaratıcı değil, zorlantılı ve genel olarak

cansız. Okulda matematikte güçlü ama sosyal ve insani bilimlerde zayıftı." Sizden Bill'in çeşitli meslek ve hobilere sahip olma olasılığını tahmin etmenizi istedik ve bu soruyu olasılık yargısının yerine benzerlik yargısı koyarak cevapladığınızı gördük. Tam olarak Bill'in muhasebeci olmasının ne kadar olası olduğunu değil, o mesleğin stereotipine ne kadar benzer olduğunu sordunuz. Şimdi cevapsız bıraktığımız bir soruya dönüyoruz: Bu yargıya nasıl vardınız?

Bill'in tanımının mesleklerin ve hobilerin stereotiplerine ne derece uygun olduğunu değerlendirmek zor değil. Bill'in cazla uğraşan insan tiplemesine bir muhasebeciden daha az benzer olduğu çok açık, bir sörfçüyle benzerliği daha da az. Örnek eşleştirmenin insanlarla ilgili yargılarda özellikle bariz olan olağanüstü çok yönlülüğünü temsil ediyor. Bill hakkında cevaplayabileceğiniz soruların bir sınırı yok. Örneğin, ıssız bir adada onunla mahsur kalmak konusunda ne hissederdiniz? Muhtemelen sağlanan kısıtlı bilgiye dayanarak bu soruya hızlı ve sezgisel bir cevap vermişsinizdir. Ancak size bir haberimiz var: Bill tanıdığımız kadarıyla olağanüstü hayatta kalma becerileri olan tecrübeli bir kâşif. Bu sizi şaşırttıysa (ve muhtemelen şaşırtmıştır) az önce tutarlılığa ulaşmada başarısız oldunuz.

Sürpriz büyük oldu çünkü bu yeni bilgi daha önce inşa ettiğiniz Bill imajıyla tutarlı değil. Şimdi Bill'in hünerinin ve hayatta kalma becerilerinin ilk tarife dâhil edildiğini düşünün. Kafanızda adamla ilgili çok farklı bir imaj oluşurdu. Belki sadece açık havada ve doğadayken canlanan bir adam. Bill'e dair genel izleniminiz daha az tutarlı ve dolayısıyla meslek ve hobi kategorilerine uydurulması daha zor olurdu ama az önce yaşadığınızdan çok daha az uyumsuzluk yaşardınız.

Çelişkili ipuçları bir tutarlılık hissine erişmeyi ve tatmin edici bir eşleşme yargısına ulaşmayı güçleştirir. Çelişkili ipuçlarının varlığı, içlerinde çok fazla gürültü bulmayı beklediğimiz karmaşık yargıların özelliğidir. Göstergelerin bazılarının olumlu, diğerle-

rinin olumsuz olduğu Gambardi problemi bu tür bir yargıydı. Karmaşık yargılara Bölüm 16'da döneceğiz. Bu bölümün geri kalanında nispeten basit yargılara, özellikle *yoğunluk ölçekleri* üstünde yapılanlara odaklanacağız.

Eşleşen Yoğunluklar

Yargıları üstünde ifade ettiğimiz ölçeklerin bazıları nitelikseldir: Meslekler, hobiler ve tıbbi teşhisler bunlara örnektir. Ölçeğin değerlerinin sıralı olmamasıyla tanımlanırlar: Kırmızı maviden ne az ne de çoktur.

Ancak pek çok yargı nicel yoğunluk ölçekleri üstünde belirlenir. Ölçü, kilo, parlaklık, sıcaklık ve ses seviyesi, maliyet ya da değer ölçümleri, olasılık ya da sıklık yargıları. Bunların hepsi niceldir. Tıpkı daha soyut ölçekler üstündeki öz güven, güç, çekicilik, öfke, korku, ölümsüzlük ya da cezaların sertliği gibi.

Bu nicel boyutların ortak ayırt edici özelliği "Hangisi daha fazla?" sorusunun aynı boyutta herhangi bir değer çifti hakkında yanıtlanabilmesidir. Kırbaç cezasının, bileğe bir fiske atmaktan daha sert bir ceza olduğunu ya da *Hamlet*'i *Godot'yu Beklerken*'den daha fazla sevdiğinizi, tıpkı Güneş'in Ay'dan daha parlak olduğunu, bir filin bir hamsterden daha ağır olduğunu ve Miami'deki ortalama hava sıcaklığının Toronto'dan daha yüksek olduğunu söyleyebildiğiniz gibi söyleyebilirsiniz.

İnsanlar bir yoğunluk ölçeğiyle bir diğeri arasında bağ kurarak, alakasız boyutlar arasında yoğunlukları eşleştirmek konusunda hatırı sayılır bir sezgisel beceriye sahiptir. Farklı şarkıcılara sevginizin büyüklüğünü, şehrinizdeki binaların yüksekliğiyle eşleştirebilirsiniz. (Örneğin, Bob Dylan'ın özellikle harika olduğunu düşünüyorsanız ona olan düşkünlüğünüzü şehrinizin en yüksek binasıyla karşılaştırabilirsiniz.) Ülkenizdeki politik anlaşmazlık seviyesini çok iyi bildiğiniz bir şehrin yaz sıcağıyla eşleştirebilirsiniz. (Hatırı

sayılır bir politik uyum varsa, bunu New York'ta sıcaklığın 20 derece civarında olduğu esintili bir yaz günüyle eşleştirebilirsiniz.) Ve sizden bir restoranla ilgili görüşünüzü, 1 ile 5 yıldız arasında değişen bir ölçek yerine, bir romanın uzunluğuyla karşılaştırarak ifade etmeniz istense, bu talep size tuhaf gelebilir ama olmayacak şey değildir. (En sevdiğiniz restoran *Savaş ve Barış* gibi olabilir.) Her durumda, tuhaf bir şekilde, ne demek istediğiniz çok açıktır.

Sıradan konuşmalarda bir ölçek için değer aralığı bağlamın bir işlevidir. "Çok para biriktirdi" yorumu, başarılı bir yatırım bankacısının emekliliği için kadeh kaldırırken ve bebek bakıcılığı yapan bir ergeni kutlarken çok farklı anlamlar taşır. *Büyük* ve küçük gibi kelimeler tamamen referans çerçevesine bağlıdır. Örneğin, "Büyük fare küçük filin hortumundan yukarı koştu", gibi bir cümleye anlam verebiliriz.

Eşleşen Öngörüler Ön Yargısı

Aşağıdaki bulmaca hem eşleştirmenin gücünü hem de onunla ilişkilendirilen sistematik bir yargı hatasını temsil ediyor.

> *Julie bir üniversitede mezun olmak üzere olan bir öğrenci. Hakkındaki bilgiyi okuduktan sonra not ortalamasını (standart 0,0 ve 4,0 ölçeğinde) tahmin edin.*
>
> *Julie dört yaşındayken akıcı bir şekilde okuyabiliyordu. Not ortalaması kaçtır?*

ABD'deki not ortalaması sistemine aşinaysanız aklınıza hemen bir sayı gelmiştir ve muhtemelen 3,7 ya da 3,8 civarındadır. Julie'nin not ortalaması hakkında bir tahminin anında akla gelmesi, az önce tarif ettiğimiz eşleştirme sürecinin yansımasıdır.

Öncelikle, Julie'nin çok erken yaşta okumaya başlamasını değerlendirdiniz. Bu değerlendirme kolaydı çünkü Julie alışıl-

madık bir yaşta okumaya başlamıştı ve bu erken gelişim Julie'yi bir ölçekte bir kategoriye yerleştirdi. Kullandığınız ölçeği tarif etmeniz gerekseydi muhtemelen en yüksek kategorisinin "olağanüstü erken gelişim" gibi bir şey olduğunu söyler ve Julie'nin bu kategoriye girmediğini (bazı çocukların daha iki yaşına bile gelmeden okuduklarını) belirtirdiniz. Julie muhtemelen bir sonraki kategoriye, "alışılmadık olmakla birlikte olağanüstü olmayan bir erken gelişim" bandına ait.

İkinci adımda Julie'yle ilgili değerlendirmenizi bir not ortalaması yargısıyla eşleştirdiniz. Bunu yaptığınızın farkında olmasanız da "alışılmadık olmakla birlikte olağanüstü olmayan" etiketine uyacak bir not ortalaması değeri aramış olmalısınız. Julie'nin hikâyesini duyunca, aklınıza, görünürde bir anda, *eşleşen bir öngörü* geldi.

Bu değerlendirme ve eşleştirme görevlerini gerçekleştirmek için gereken hesapları yavaş ve düşünerek yapmak biraz zaman alırdı ama hızlı, Sistem 1 düşünmede, yargı hızlı ve çaba harcanmadan gerçekleştirilir. Burada Julie'nin not ortalamasını tahmin etmekle ilgili anlattığımız hikâye direkt olarak gözlenemeyen zihinsel olayların karmaşık, çok kademeli sekanslarını içerir. Eşleştirmenin zihinsel mekanizmasının özgüllüğü psikolojide sıra dışıdır ama kanıtı alışılmadık ölçüde katidir. Birçok benzer deneyden, aşağıdaki iki sorunun farklı insan gruplarına sorulduklarında tamamen aynı sayıları vereceklerinden emin olabiliriz.

- *Julie'nin sınıfının yüzde kaçı ondan daha erken yaşta okumaya başladı?*
- *Julie'nin sınıfının yüzde kaçının not ortalaması ondan yüksektir?*

İlki tek başına baş edilebilir bir soru: Basitçe size Julie hakkında verilen kanıtı değerlendirmenizi istiyor. Uzak bir tahmin isteyen

ikinci soru kesinlikle daha zor ama onu ilk soruyu cevaplayarak cevaplamak sezgisel olarak cazip gelecektir.

Julie'yle ilgili sorduğumuz iki soru, geçerlilik yanılsamasını ele alırken evrensel olarak kafa karıştırıcı olarak tanımladığımız iki soruya benzer. Julie'yle ilgili ilk soru sizden onun vakasıyla ilgili bilginizin "yoğunluğunu" değerlendirmenizi istiyor. İkinci soru ise bir tahminin yoğunluğunu soruyor. Ve ikisinin birbirinden ayırt edilmesinin hâlâ zor olduğundan şüphe duyuyoruz.

Julie'nin not ortalamasının sezgisel tahmini Bölüm 13'te tanımladığımız psikolojik vakanın bir örneğidir: zor bir sorunun yerine kolay bir sorunun getirilmesi. Sistem 1'iniz zor bir tahmin sorusunu çok daha kolay bir soruyu cevaplayarak kolaylaştırır: Julie'nin dört yaşında okumaya başlama başarısı ne kadar etkileyiciydi? Yılla ölçülen okuma yaşından notlarla ölçülen not ortalamasına doğrudan hareket etmek için ekstra bir eşleştirme adımı gerekli.

Elbette yerine koyma ancak eldeki bilgi alakalıysa gerçekleşir. Julie hakkında bütün bildiğiniz hızlı bir koşucu ya da vasat bir dansçı olması olsaydı, hiç bilginiz olmazdı. Ancak zekânın makul bir göstergesi olarak yorumlanabilecek her olgunun kabul edilebilir bir ikame olması mümkündür.

Bir sorunun yerine başka bir soru koymak, iki sorunun doğru cevaplarının farklı olduğu durumlarda kaçınılmaz olarak hatalara yol açar. Okuma yaşını not ortalamasının yerine koymak, makul gibi görünse de açıkça absürttür. Nedenini anlamak için Julie'nin dört yaşından beri yaşanabilecek olayları düşünün. Korkunç bir kaza geçirmiş olabilir. Anne ve babası travmatik bir şekilde boşanmış olabilir. Onu çok etkileyen ilham verici bir öğretmenle tanışmış olabilir. Hamile kalmış olabilir. Bütün bu olaylar ve daha pek çoğu üniversitedeki performansını etkilemiş olabilir.

Eşleşen öngörü sadece okuma yaşının erkenliği ve üniversite not ortalaması kusursuz bir şekilde bağıntılı ise doğrulanabilir; ki burada öyle olmadığı çok açık. Öte yandan Julie'nin okuma yaşıyla

ilgili bilgiyi tamamen yok saymak da bir hata olurdu çünkü okuma yaşı bize birtakım alakalı bilgiler sağlıyor. Optimal tahmin, kusursuz bilgi ve sıfır bilginin uç noktaları arasında bir yerde olmalı.

Bir vaka hakkında ait olduğu kategori dışında özel bir bilginiz olmadığında, ne biliyorsunuzdur? Bu sorunun cevabı vakaya dış bakış olarak adlandırdığımız şeydir. Julie'nin not ortalamasını tahmin etmemiz istenseydi ama bize hakkında hiçbir bilgi verilmeseydi, büyük olasılıkla genel ortalamayı tahmin ederdik, belki de 3,2 derdik. Buna dış bakış öngörüsü denir. Julie'nin not ortalamasının en iyi tahmini dolayısıyla 3,2'den yüksek, 3,8'den düşük olmalıdır. Tahminin tam konumu bilginin kestirimci değerine bağlıdır; okuma yaşına not ortalamasının bir kestiricisi olarak ne kadar çok güvenirseniz, tahmin o kadar yükselir. Julie'nin örneğinde, bilgi elbette hayli zayıf ve buna göre en akla yatkın tahmin genel not ortalamasına yakın olacaktır. Ama eşleşen öngörüler hatasını düzeltmenin teknik ama hayli kolay bir yolu var: Ek C'de detaylandıracağız.

İstatistiksel açıdan absürt öngörülere yol açsalar da kanıtla eşleşen tahminlere karşı koymak zordur. Satış müdürleri sıklıkla geçen yıl satış ekibinden daha başarılı olan satışçının yine hepsinden daha iyi performans sağlayacağını varsayarlar. Kıdemli üst düzey yöneticiler bazen son derece yetenekli bir adayla tanışır ve işe yeni alınacak bu elemanın örgütün en tepesine yükseleceğini hayal ederler. Yapımcılar rutin olarak, bir önceki filmi hit olan bir yönetmenin bir sonraki filminin de en az onun kadar başarılı olmasını beklerler.

Bu eşleşen öngörü örneklerinin hayal kırıklığıyla son bulma ihtimali bulmaması ihtimalinden daha yüksektir. Öte yandan, durumun en kötü olduğu zamanda yapılan öngörü eşleştirmeleri aşırı olumsuz olma eğilimindedirler. Kanıtla eşleşen sezgisel öngörüler iyimser olduklarında da kötümser olduklarında da aşırıdırlar. (Bu tür tahmin hataları için teknik terim *regresif olmayandır* çünkü

ortalamaya gerileme adı verilen bir istatistiki fenomeni hesaba katmazlar.)

Öte yandan ikamenin ve eşleştirmenin tahminleri her zaman yönetmediği de kayda geçirilmelidir. İki sistemin dilinde, sezgisel Sistem 1 sorunlara daha doğarken hızlı çağrışımsal çözümler önerir ama bu sezgiler inanca dönüşmeden daha fazla düşünmeye dayalı Sistem 2 tarafından desteklenmelidir. Eşleşen öngörüler bazen daha karmaşık cevapların lehine reddedilir. Örneğin, insanlar tahminleri aleyhteki kanıtlarla eşleştirmeye, lehteki kanıtlarla eşleştirmekten çok daha isteksizdirler. Julie okumaya geç başlamış olsaydı daha düşük bir üniversite performansıyla eşleşen bir öngörüde bulunurken tereddüt edeceğinizi düşünüyoruz. Lehteki ve aleyhteki tahminler arasındaki asitmetri daha fala bilgiye erişilebildiği zaman kaybolur.

Dış bakışı her tür sezgisel öngörü için bir düzeltici olarak sunuyoruz. Michael Gambardi'nin gelecekteki olasılıklarıyla ilgili tartışmamızda, Michael'in gelecekteki başarı olasılığıyla ilgili yargınızı ilgili baz oranına (işe yeni başlayan CEO'ların iki yıllık başarı oranı) bağlamanızı tavsiye ederiz. Julie'nin not ortalaması gibi nicel öngörülerde, dış bakışı benimsemek tahmininizi ortalama sonuca bağlamanız anlamına gelir. Dış bakış sadece erişilebilir bilginin tahminin tam bir güvenle yapılmasını desteklediği çok kolay sorularda yok sayılabilir. Ciddi yargı gerekli olduğunda, dış bakış çözümün bir parçası olmalıdır.

Eşleştirmede Gürültü: Mutlak Yargının Sınırlamaları

Yoğunluk ölçeklerinde kategorileri ayırt etmedeki becerimizin kısıtlı olması, eşleştirme işleminin isabetliliğini sınırlar. *Büyük* ya da *zengin* gibi sözcükler ölçü ya da zenginlik boyutunda belli bir değer aralığına aynı etiketi verir. Bu potansiyel olarak önemli bir gürültü kaynağıdır.

Emekli olan yatırım bankacısı *zengin* etiketini elbette hak eder ama ne kadar zengindir? Aralarından seçim yapabileceğimiz sıfatlar çok çeşitli: *iyi durumda, varlıklı, rahat, zengin, süper zengin vs.* Size birtakım bireylerin varlık durumları hakkında detaylı tanımlar verilseydi ve her birine bir sıfat vermeniz gerekseydi, vakalar arasında detaylı karşılaştırmalara başvurmadan, kaç ayrı kategori oluşturabilirdiniz?

Bir yoğunluk ölçeği üzerinde ayırt edebileceğimiz kategorilerin sayısı 1956'da yayımlanan ve psikoloji alanında bütün zamanların klasiklerinden biri sayılan makalenin isminde belirtilmiştir: "Sihirli Sayı Yedi, Artı Eksi İki." Bu sınırın ötesinde insanlar hata yapmaya başlamaya meylederler, örneğin kafa kafaya bir karşılaştırmada aslında B'ye A'dan daha yüksek bir puan verecekken, B'den daha yüksek bir kategoriye A vermek gibi.

5 ve 10 santim arasında farklı uzunluklarda 4 çizgi düşünün. Her çizgi bir öncekiyle aynı uzunlukta olsun. Size her defasında bir çizgi gösterilsin ve 1 en kısa çizgi, 4 en uzun çizgi olmak üzere, 1 ile 4 arasında bir sayıyla tanımlamanız istensin. Basit bir görev. Şimdi farklı uzunluklarda 5 çizgi gösterildiğini ve aynı görevi 1 ile 5 arasında sayılar vererek tekrarlamanız gerektiğini düşünün. Ne zaman hata yapmaya başlarsınız? Çizgiler sihirli sayı yediye ulaştığında. Şaşırtıcı bir şekilde, bu sayı çizgi uzunluklarının aralığına çok bağlı değildir. Yani çizgilerin uzunluğu 5 ve 10 yerine 5 ve 15 santim arasında değişseydi bile, yedi çizginin ötesinde hatalar yapmaya başlardınız. Aynı görev farklı yüksekliklerde sesler ya da farklı parlaklığa sahip ışıklarla yapılsaydı, hemen hemen aynı sonuç elde edilirdi. İnsanların bir boyutta uyarıcılara farklı etiketler verme becerisinin hakiki bir sınırı vardır ve o sınır, yedi etiket civarındadır.

Ayırt etme gücümüzün sınırı önemlidir çünkü yoğunluk boyutları arasında değer eşleştirme becerimiz bu boyutlar üstünde değer atama becerimizden daha iyi olamaz. Eşleştirme işlemi hızlı,

Sistem 1 düşünmenin çok yönlü bir aracı ve pek çok sezgisel yargının özüdür ama hamdır.

Sihirli sayı mutlak bir sınır değildir. İnsanlar hiyerarşik sınıflandırmayla daha ince ayrımlar yapmak üzere eğitilebilirler. Örneğin, multi milyonerler arasında çok sayıda varlık kategorisi belirleyebiliriz ve kendileri de sertliklerine göre sıralanan yargıçlar çok sayıda suç kategorisinde ciddiyet derecelerini ayırabilir. Ancak bu rafine etme sürecinin işlemesi için kategoriler önceden var olmalı ve sınırları net olmalıdır. Bir çizgi grubuna etiket tayin ederken, uzun çizgileri kısalardan ayırmaya ve onlara iki ayrı kategori gibi davranmaya karar veremezsiniz. Siz hızlı düşünme modundayken, sınıflandırma istemli kontrol altında değildir.

Sıfat ölçeklerinin sınırlı ayrım gücünü aşmanın bir yolu var: etiketler yerine, karşılaştırmalar kullanmak. Vakaları karşılaştırma becerimiz onları bir ölçeğe yerleştirme becerimizden çok daha iyidir.

Büyük bir restoran ya da şarkıcı kümesini değerlendirmek için yirmi noktalı bir kalite ölçeği kullanmanız istendiğinde, ne yapardınız? Bir beş yıldız ölçeğinin altından kalkmak kolay olurdu ama yirmi noktalı bir ölçekte kusursuz güvenilirliği koruyamazdınız. (Joe's Pizza üç yıldızlık bir restoran ama on bire mi yoksa on ikiye mi karşılık geliyor?) Bu sorunun çözümü zaman alıcı olmakla birlikte, basit. Önce restoranları ya da şarkıcıları beş kategoriye ayırmak için beş noktalı ölçekte puanlarsınız. Daha sonra her kategorideki örnekleri sıraya koyarsınız ve genelde bunu sadece birkaç bağla yapabilirsiniz. Onları aynı kategoriye koysanız bile, muhtemelen Joe's Pizza'yı Fred's Burgers'a ya da Taylor Swift'i Bob Dylan'a tercih edip etmediğinizi biliyorsunuzdur. İşi basitleştirmek için şimdi beş kategorinin her birinin içinde dört seviye ayırabilirsiniz. Muhtemelen en sevmediğiniz şarkıcıları bile hor görme seviyelerinize göre ayırabilirsiniz.

Bu egzersizin psikolojisi çok açıktır. Yargı nesneleri arasında aleni karşılaştırmalar tek seferde değerlendirilen nesnelerin pu-

anlamalarından çok daha ince ayrımları destekler. Çizgi boyu yargıları benzer bir hikâye anlatır: Peş peşe gösterilen çizgilerin uzunluğunu karşılaştırma beceriniz uzunlukları sınıflandırma becerinizden daha iyidir ve aynı anda görülebilen çizgileri karşılaştırırken daha da isabetli olursunuz.

Karşılaştırmalı yargıların avantajı pek çok alana uyar. İnsanların servetleri konusunda kabaca bir fikriniz varsa, aynı aralıktaki insanları karşılaştırırken, servetlerini bireysel olarak etiketlerken çıkaracağınızdan daha iyi iş çıkarırsınız. Ödev puanlarken, en iyiden en kötüye sıraladığınızda, ödevleri tek tek okuyup puanladığınız zamana göre daha net olabilirsiniz. Karşılaştırmalı ya da göreceli yargılar, kategorik ya da mutlak yargılara göre daha hassastır. Bu örneklerin gösterdiği gibi, ayrıca daha fazla çaba gerektirir ve zaman alırlar.

Nesneleri açıkça karşılaştırmalı ölçeklerde ayrı ayrı puanlamak karşılaştırmalı yargının bazı faydalarını korur. Bazı bağlamlarda, özellikle eğitimde, adayların kabul edilmesi ya da terfi ettirilmesi önerileri genellikle tavsiyede bulunan kişinin adayı "ders verdiğiniz öğrenciler" ya da "aynı tecrübe seviyesine sahip programcılar" gibi belirli bir nüfusun "ilk %5"i ya da "ilk %20"si içine yerleştirmesini gerektirir. Bu puanlamalar oldukları gibi alınmayı nadiren hak ederler çünkü tavsiye sahiplerini ölçeği düzgün kullanmaktan sorumlu tutmanın bir yolu yoktur. Sorumluluk bazı bağlamlarda mümkün olur: Müdürler çalışanları puanladığında ya da analizciler yatırımları değerlendirirken, vakaların %90'ını "ilk %20" sınıfına sokan bir kişi teşhis edilip düzeltilebilir. Karşılaştırmalı yargıların kullanımı Kısım 5'te ele alacağımız gürültü çarelerinden biridir.

Pek çok yargı görevi bireysel örnekleri bir ölçek üzerinde (örneğin yedi noktalık bir anlaşma ölçeğinde) bir kategoriyle eşleştirmeyi ya da sıralı bir sıfat grubunu (örneğin, olayların olasılığını puanlarken "olasılıksız" ya da "son derece olasılıksız" gibi) kullanmayı gerektirir. Bu eşleme türü, ham olduğu için gürültülüdür.

Bireyler yargının özü üzerinde hemfikir olsalar bile, etiketlerin yorumu konusunda farklılık sergileyebilirler. Aleni karşılaştırmalı yargıyı gerektiren bir prosedürün gürültüyü azaltması olasıdır. Bir sonraki bölümde yanlış ölçek kullanmanın gürültüyü nasıl artırdığını göreceğiz.

Eşleştirmeden Bahsederken

"İkimiz de bu film çok iyi diyoruz ama sen benden daha az keyif almış gibi görünüyorsun. Aynı sözcükleri kullanıyoruz ama aynı ölçeği kullanıyor muyuz?"

"Bu dizinin 2. sezonu 1. sezon kadar izlenesi olur diye düşünüyorduk. Ama eşleşen bir öngörüde bulunduk ve yanlış çıktı."

"Bu ödevleri notlarken tutarlı kalmak çok zor. Acaba sıraya koymayı mı denesen?"

BÖLÜM 15

Ölçekler

Kendinizi özel bir hukuk davasında jüri üyesi olarak düşünün. Aşağıda özetlenen kanıtları dinlediniz ve sizden bu konuda bazı yargılarda bulunmanız istenecek.

Joan Glover – General Asistance davası

Altı yaşında bir çocuk olan Joan Glover reçetesiz bir alerji ilacı olan Allerfree'den çok sayıda yutar ve kapsamlı bir bakım almak için hastanede kalması gerekir. Aşırı doz solunum sistemini zayıf düşürdüğü için, hayatının geri kalanında astım ve amfizem gibi solunumla ilgili hastalıklara daha açık olacaktır. Allerfree'nin şişesinde çocuk güvenliği açısından yetersiz bir kapak kullanılmaktadır.

Allerfree'nin üreticisi General Asistance çeşitli reçetesiz ilaçlar üreten büyük bir firmadır (yıllık kârı 100 ile 200 milyon arasında değişiyor). Federal yönetmelik bütün ilaç şişelerinde çocuk korumalı güvenlik kapakları kullanılmasını şart koşmaktadır. General Asistance sorun çıkarma oranı sektördeki diğer örneklerinden daha yüksek olan bir çocuk korumalı güvenlik kapağı kullanarak bu

yönetmeliğin maksadını sistematik bir şekilde yok saymaktadır. Şirkete ait bir iç yazışmada "bu aptal, gereksiz federal yönetmeliğin paralarını ziyan etmekten başka işe yaramadığı" ve cezalandırılma riskinin düşük olduğu belirtilmiştir. Aynı belgede, her durumda, "yönetmeliği çiğneme cezalarının son derece hafif olduğu, şirketten sadece gelecekte güvenlik kapaklarının iyileştirilmesinin talep edileceği" de eklenmiştir. Gıda ve İlaç İdaresi'nin bir görevlisi tarafından uyarılmalarına rağmen, şirket düzeltici bir faaliyette bulunmaya gerek görmemiştir.

Şimdi sizden üç yargıda bulunmanızı isteyeceğiz. Lütfen cevaplarınızı seçmek için hızınızı yeterli ölçüde düşürün.

Öfke:

Sanığın eylemleriyle ilgili görüşünüzü aşağıdakilerden hangisi en iyi ifade eder? (Lütfen cevabınızı daire içine alın.)

Tamamen kabul edilebilir		*İtiraz edilebilir*		*Şoke edici*		*Kesinlikle dehşet verici*
0	1	2	3	4	5	6

Cezalandırma niyeti:

Maddi tazminat ödemeye ek olarak, sanık ne kadar cezalandırılmalı? (Lütfen uygun ceza seviyesi için görüşünüzü en iyi ifade eden sayıyı daire içine alın.)

Ceza olmasın		*Hafif ceza*		*Ağır ceza*		*Son derece ağır ceza*
0	1	2	3	4	5	6

Tazminat: Maddi tazminata ek olarak, sanıktan ceza olarak ve gelecekte sanığı ve diğerlerini benzer eylemlerden caydırması amacıyla ne kadar *cezalandırıcı* tazminat talep edilmeli (edilmeliyse)? (Lütfen cevabınızı aşağıdaki boş kısma yazın.)

$.....

Joan Glover'ın hikâyesi ikimizin (yani Kahneman ve Sunstein'ın dostumuz ve çalışma arkadaşımız David Schkade ile birlikte) 1998'de raporladığımız bir çalışmada kullanılan bir örneğin kısaltılmış versiyonudur. Bu çalışmayı bu bölümde bir miktar detaylandıracağız. Çalışmanın içerdiği görevlerden birini tecrübe etmenizi istedik çünkü bu vakayı artık bu kitabın pek çok temasını içeren gürültü denetiminin eğitici bir örneği olarak görüyoruz.

Bu bölüm *yanıt ölçeğinin* yaygın bir gürültü kaynağı olarak rolüne odaklanacak. İnsanlar yargılarında özde anlaşmazlık yaşadıkları için değil, ölçeği farklı şekillerde kullandıkları için de farklılık sergileyebilirler. Bir çalışanın performansını puanlarken 0 ile 6 arası bir ölçekte, performansının 4 olduğunu söyleyebilirsiniz; sizin bakış açınıza göre bayağı iyi bir puandır. Bir başkası aynı ölçekte çalışanın performansının 3'lük olduğunu söyleyebilir; onun bakış açısıyla bu da bayağı iyidir. Ölçeklerin sözlü ifadesindeki muğlaklık genel bir sorundur. "Şüpheye yer bırakmayacak şekilde", "açık ve ikna edici kanıt", "çarpıcı performans" ve "olması pek mümkün olmayan" gibi muğlak ifadelerden doğan iletişim zorlukları hakkında pek çok araştırma yapıldı. Bu tür kelime gruplarıyla ifade edilen yargılar kaçınılmaz olarak gürültülüdürler çünkü hem konuşmacılar hem dinleyiciler tarafından farklı yorumlanırlar.

Joan Glover vakasının konu olduğu çalışmada, muğlak bir ölçeğin ciddi sonuçlara yol açtığı bir durumda etkilerini gözlemledik. Çalışmanın konusu jürinin cezalandırıcı tazminat kararlarındaki gürültüydü. Joan Glover vakasının üçüncü sorusundan anlayabileceğiniz gibi, ABD'de (ve başka ülkelerde) hukuk, özel davalarda jürilerin eylemleri özellikle korkunç olan sanıklara cezalandırıcı tazminat dayatmasına olanak verir. Cezalandırıcı tazminatlar, zarar gören kişinin zararını gidermeyi hedef alan maddi tazminatlara ek olarak talep edilir. Glover örneğinde olduğu gibi, bir ürün yaralanmaya neden olduysa ve davacı şirketi dava ettiyse ve haklı çıktıysa, tıbbi faturalarını ödemek ve kaybedilen geliri

telafi etmek için para alır. Ama aynı zamanda, davalıya ve benzer şirketlere bir uyarı mahiyetinde cezalandırıcı tazminat da alabilir. Bu vakada General Asistance bariz bir şekilde kınanacak bir davranış sergiledi; bu davranış bir jürinin akla yatkın bir şekilde cezalandırıcı tazminat talep edebileceği eylemler sınıfına giriyor.

Cezalandırıcı tazminat kurumuyla ilgili önemli bir endişe, öngörülemezliğidir. Aynı kabahat çok mütevazı ve yüklü arasında değişen çok farklı tazminatlarla cezalandırılabilir. Bu kitabın terminolojisini kullanarak sistemin gürültülü olduğunu söyleyebiliriz. Cezalandırıcı tazminat talepleri genellikle geri çevrilir ve verildikleri zaman, miktar genellikle maddi tazminata büyük bir ek getirmez. Bununla birlikte çarpıcı istisnalar da olur ve bazen jürilerin karara bağladığı çok büyük meblağlar şaşırtıcı ve keyfî görünür. En sık bahsi geçen örneklerden biri, davacının sıfır BMW'sinin tekrar boyanmış olduğu bilgisinin gizlenmesi nedeniyle bir araba bayisine verilen 4 milyon dolarlık tazminat cezasıdır.

Cezalandırıcı tazminatlar çalışmamızda, 899 katılımcıdan Joan Glover'ın vakasını ve dokuz benzer vakayı değerlendirmeleri istendi. Vakaların hepsinde belli miktarda zarar görmüş ve sorumlu olduğu iddia edilen şirketi dava etmiş davacılar vardı. Sizin aksinize, katılımcılar on vakanın hepsinde üç sorudan (öfke, cezai niyet ya da dolar olarak ceza miktarı) sadece birini cevapladılar. Katılımcılar daha sonra daha küçük gruplara ayrıldılar ve her gruba her davanın bir versiyonu verildi. Farklı versiyonlarda davacının aldığı zarar ve davalı şirketin geliri farklılık gösteriyordu. Toplam yirmi sekiz senaryo vardı. Amaçlarımız cezalandırıcı tazminatların psikolojisi hakkında bir teoriyi test etmek ve bu yasal kurumda parasal ölçeğin (burada dolar) başlıca gürültü kaynaklarından biri olarak rolünü soruşturmaktı.

Öfke Hipotezi

Adil bir cezanın nasıl belirlenmesi gerektiği konusu yüzyıllardır filozoflar ve hukuk bilimciler tarafından tartışılmaktadır. Ancak bizim hipotezimiz, filozofların zor bulduğu sorunun görevi zor soruyu kolay bir soruyla değiştirerek basitleştiren sıradan insanlar için hayli basit olduğuydu. Size General Asistance'ın ne kadar cezaya çarptırılması gerektiği sorulduğunda, önce ve hızla cevaplanan kolay soru şudur: "Ne kadar açım?" Daha sonra planlanan cezanın yoğunluğu, öfkenin yoğunluğuyla eşleştirilir.

Bu **öfke** hipotezini test etmek için farklı katılımcı gruplarından ya cezai niyet sorusunu ya da öfke sorusunu cevaplamalarını istedik. Daha sonra çalışmada kullanılan yirmi sekiz senaryo için iki sorudan elde edilen ortalama puanları karşılaştırdık. İkame fikrinden beklendiği gibi, öfke ve cezai niyetin ortalama puanları arasındaki korelasyon kusursuza yakın bir 0,98 (YU = %94) idi. Bu korelasyon öfke hipotezini destekliyor. Öfke duygusu cezai niyetin birincil belirleyicisidir.

Öfke cezai niyetin ana yönlendiricisidir ama yalnız değildir. Joan'ın hikâyesinde, cezai niyeti puanlarken, öfkeyi puanlarken olduğundan daha fazla dikkat çeken bir şey fark ettiniz mi? Fark ettiyseniz, bunun Joan'ın gördüğü zarar olduğunu tahmin ediyoruz. Bir davranışın korkunç olup olmadığını sonuçlarını bilmeden anlayabilirsiniz. Bu örnekte, General Asistance'ın davranışı hiç şüphesiz korkunçtu. Aksine, cezai niyetle ilgili sezgilerin göze göz prensibinde kabaca ifade bulan cezalandırıcı bir yönü vardı. Cezalandırma arzusu, cinayete teşebbüsle cinayetin kanun ve yargıçlar nezdinde neden farklı muamele gördüğünü ve hedefi ıskalayacak kadar şanslı olan muhtemel katilin neden daha az ceza aldığını açıklar.

Maruz kalınan zararın cezai niyette fark yaratırken, öfkede de yaratıp yaratmadığını öğrenmek için, farklı cevap gruplarına Joan Glover vakasının ve diğer birkaç vakanın "ağır hasar" ve

"hafif hasar" versiyonlarını gösterdik. Ağır hasar versiyonu sizin gördüğünüzdü. Hafif hasar versiyonunda, Joan "bir süre hastanede kalmak zorundaydı ve her tür hap karşısında derin bir travma içindeydi. Anne ve babası ona vitamin, aspirin, soğuk algınlığı ilaçları gibi faydalı ilaçlar vermeye çalıştığında bile, kontrolsüz bir şekilde ağlıyor ve korktuğunu söylüyordu". Bu örnek, çocuk için travmatik bir tecrübeyi tarif eder ama okuduğunuz ilk versiyondaki uzun vadeli tıbbi hasardan daha düşük bir hasar düzeyi söz konusudur. Beklendiği gibi, ortalama öfke puanlamaları ağır hasar örneğinde (4,24) ve hafif hasar örneğinde (4,19) neredeyse aynıydı. Öfkede sadece davalının davranışı önemlidir; sonuçları değil. Aksine, cezai niyet ortalamaları ağır hasar için 4,93 ve hafif hasar için 4,65 idi. Küçük ama istatistiksel açıdan güvenilir bir fark. Ağır hasar versiyonu için medyan ceza iki milyon dolar, daha hafif versiyon içinse bir milyon dolardı. Daha pek çok vakada benzer sonuçlar elde edildi.

Bu bulgular yargı sürecinin kilit bir özelliğini vurguluyor: kanıtın farklı yönlerini tartmada yargı görevinin incelikli etkisi. Cezai niyeti ve öfkeyi puanlayan katılımcılar felsefi bir mesele olan, "Adalet cezalandırıcı olmalı mı?" meselesini ilgilendiren bir duruş sergilediklerinin farkında değildiler. Vakanın çeşitli özelliklerine ağırlıklar atfettiklerinin bile farkında değildiler. Bununla birlikte, öfkeyi puanlarken, zarara sıfıra yakın bir ağırlık atfederken, iş cezayı belirlemeye geldiğinde, aynı faktöre hatırı sayılır bir ağırlık verdiler. Katılımcıların hikâyenin sadece bir versiyonunu gördüklerini, daha kötü zarara daha yüksek ceza vermelerinin aleni bir karşılaştırma olmadığını hatırlayalım. Bu, iki koşulda otomatik bir eşleştirme operasyonunun sonucuydu. Katılımcıların cevapları yavaş düşünmeden çok, hızlı düşünmeye dayanıyordu.

Gürültülü Ölçekler

Çalışmanın ikinci hedefi cezalandırıcı tazminatlarının neden gürültülü olduğunu bulmaktı. Hipotezimiz jüri üyelerinin davalının ne kadar ağır cezalandırılması gerektiği konusunda genellikle hemfikir oldukları ama cezai niyetlerini dolar ölçeğine çevirirken çok büyük farklılıklar sergiledikleriydi.

Çalışmanın tasarımı aynı vakaların yargılarındaki gürültü miktarını üç ölçekte karşılaştırmamıza izin veriyor: öfke, cezai niyet ve dolar olarak ceza tazminatı. Gürültüyü ölçmek için Bölüm 6'da federal yargıçların gürültü sonuçlarını analiz etmek için kullanılan yöntemi uyguluyoruz. O analizde de olduğu gibi, bir vakanın bireysel yargılarının ortalamasının ön yargısız, adil bir değer olarak ele alınabileceğini varsayıyoruz. (Bu analizin amaçları doğrultusunda bir varsayımdır, yanlış olabileceğini vurgulamak isteriz.) İdeal bir dünyada belirli bir ölçek kullanan bütün jüri üyeleri her vakada yargıları konusunda hemfikir olurdu. Ortalama yargıdan her sapma, hata olarak görülür ve bu hatalar sistem gürültüsünün kaynağıdır.

Bölüm 6'da da dikkat çektiğimiz gibi, sistem gürültüsü yatay gürültü ve örüntü gürültüsü olarak parçalara ayrılabilir. Burada, yatay gürültü jüri üyeleri arasında genel olarak ne kadar katı oldukları konusundaki farklılıktır. Örüntü gürültüsü ise belli bir jüri üyesinin kendi ortalamasına göre belirli vakalara verdiği tepkilerdeki farklılıktır. Bu nedenle, yargıların toplam varyansını üç ögeye ayırabiliriz.

Yargıların Varyansı =
Adil Cezaların Varyansı + (Yatay Gürültü)^2 + (Örüntü Gürültüsü)^2

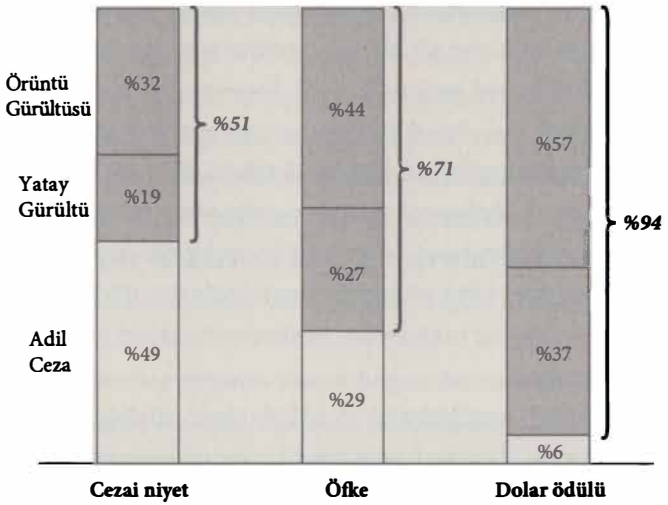
Yargıların farklılığını üç terime ayıran bu analiz, öfke, cezai niyet ve para cezası miktarı yargılarından ayrı olarak yürütüldü.

Şekil 13 sonuçları gösteriyor. En az gürültülü ölçek, sistem gürültüsünün farklılığın %51'inden sorumlu olduğu cezai niyet: Bu, adalet kadar gürültü var demek. Öfke ölçeği bariz bir şekilde daha gürültülü: %71 gürültü. Para cezası miktarı ölçeği ise açık ara en kötüsü: Yargılardaki farklılığın %94'ü gürültü!

Farklar oldukça çarpıcı çünkü bu üç ölçek, içerikleri açısından neredeyse aynılar. Daha önce adil öfke ve cezai niyet değerlerinin, öfke hipotezinin öne sürdüğü gibi, neredeyse kusursuz bir şekilde bağıntılı olduğunu görmüştük. Cezai niyet ve para cezası cevaplarının puanlamaları net bir şekilde aynı soruyu –General Asistance'ın ne kadar ağır bir ceza alması gerektiğini– farklı birimlerle yanıtlar. Şekil 13'te gördüğümüz büyük farkları nasıl açıklayabiliriz?

Muhtemelen, öfkenin çok net bir ölçek olmadığında hemfikir olabiliriz. Doğru, "tamamen kabul edilebilir" davranış diye bir şey var ama General Asistance'a ya da diğer davalılara ne kadar kızacağınızın bir sınırı var ve o sınır hayli müphem. Bir davranışın "kesinlikle dehşet verici" olması ne anlama geliyor? Ölçeğin yukarı ucunda netlik olmaması gürültüyü kaçınılmaz kılıyor.

Cezai niyet ise daha belirlidir. "Ağır ceza", "kesinlikle dehşet verici"den daha nettir çünkü "son derece ağır ceza" kanunda tanımlanan maksimumla sınırlıdır. Suçluyu cezaya boğmak isteyebilirsiniz ama örneğin, General Asistance'ın CEO'sunu ve bütün üst düzey yönetim kadrosunu öldürmeyi öneremezsiniz. (Umarız.) Cezai niyet ölçeği, üst sınırının daha net belirlenmiş olması sebebiyle daha az muğlaktır. Tahmin edeceğimiz gibi, aynı zamanda daha az gürültülüdür.



ŞEKİL 13: *Yargı farklılığının ögeleri*

Öfke ve cezai niyet benzer puanlama ölçeklerinde ölçüldüler ve sözel etiketlerle üç aşağı beş yukarı net olarak tanımlandılar. Dolar ölçeği çok daha fazla sorunlu, farklı bir aileye aittir.

Dolarlar ve Çapalar

Akademik makalemizin başlığı merkezî mesajını ifade ediyor: "Ortak Öfke ve Dengesiz Cezalar: Cezalandırıcı Tazminatların Psikolojisi." Deneydeki jüri üyelerimizin cezai niyet puanları arasında büyük ölçüde görüş birliği vardı; puanlamalar çoğunlukla öfkeyle açıklanabiliyordu. Öte yandan, dolar ölçüsü mahkeme salonunu yakından taklit ediyordu ve kabul edilemeyecek kadar gürültülüydü.

Nedeni hiç de gizemli değil. Joan Glover vakasında belirli bir ceza tazminatı miktarı ürettiyseniz bir sayı tercihinizin özünde keyfî olduğu hissini yaşamışsınızdır. Keyfîlik hissi, önemli bir bilgi aktarır: Size başka insanların çok değişken keyfî kararlar

alacaklarını ve yargılarının çok gürültülü olacağını söyler. Bu dolar olarak ceza miktarının ait olduğu ölçekler ailesinin bir özelliğidir.

Efsane Harvard psikoloğu S. S. Stevens insanların, şaşırtıcı bir şekilde pek çok öznel tecrübe ve tavırda yoğunluk *oranları* hakkında güçlü sezgiler paylaştığını keşfetti. Bir ışığı bir diğerinin "iki katı parlak" görünecek şekilde ayarlayabiliyor ve on aylık bir hapis cezasının duygusal öneminin bir aylık bir hapis cezasının neredeyse on katı kötü olmadığı konusunda hemfikir olabiliyorlardı. Stevens bu tür sezgilerden beslenen ölçekleri *oran ölçekleri* olarak adlandırdı.

Parayla ilgili sezgilerimizin oranlarla ifade edildiğini, "Sara %60 zam aldı!" ya da "Zengin komşumuz bir gecede servetinin yarısını kaybetti." gibi ifadeleri ne kadar kolay anladığımızdan anlayabilirsiniz. Cezalandırıcı tazminatların dolar ölçeği, cezalandırma niyetinin ölçümü için bir orantı ölçeğidir. Diğer orantı ölçekleri gibi, anlamlı bir sıfırı (sıfır dolar) vardır ve tepede sınırsızdır.

Stevens bir onay ölçeğinin (dolar ölçeği gibi) tekil bir ara çapayla (jargondaki terimi modül) belli bir aralığa sabitlenebileceğini keşfetti. Laboratuvarında gözlemcileri belirli bir parlaklıktaki bir ışığa maruz bırakıp "o ışığın parlaklığına 10 (ya da 50 veya 200) diyerek başka parlaklıklara buna göre sayı biçme" görevi verdi. Beklendiği gibi, o gözlemcilerin farklı parlaklıklardaki ışıklara verdiği sayılar benimsemelerinin istendiği keyfî çapayla orantılıydı. 200 sayısını çapa alan bir gözlemci, çapanın 10 olması hâlinde yapacağından 20 kat yüksek yargılarda bulundu. Gözlemcinin yargılarının standart sapması da çapayla orantılı oluyordu.

Bölüm 13'te insanların bir nesneye para ödeme istekliliğinin onlara önce Sosyal Güvenlik numaralarının son iki hanesini ödemeye razı olup olmayacaklarının sorulmasından güçlü olarak etkilendiği keyifli bir çapalama örneği verdik. Daha çarpıcı bir sonuç, ilk çapanın başka nesnelerden oluşan koskoca bir listeye ödemeye istekli olacakları rakamı da ekliyor olmasıydı. Kablosuz

bir izleme topuna büyük bir meblağ ödemeye ikna edilen katılımcılar kablosuz bir klavyeye de paralel şekilde yüksek bir miktar ödemeyi kabul ettiler. Anlaşılıyor ki insanlar karşılaştırılabilir malların *göreceli* değerine mutlak değerlerine olduğundan çok daha hassas yaklaşıyorlar. Çalışmanın yazarları tekil çapanın ısrarcı etkisini "tutarlı keyfiyet" olarak adlandırdılar."

Joan Glover vakasındaki keyfî çapanın etkisini anlamak için, bu bölümün en başındaki metnin aşağıdaki bilgiyi içerdiğini varsayın.

> *Başka bir ilaç şirketini içeren benzer bir vakada, kurban olan küçük kız orta düzeyde psikolojik travma yaşadı (daha önce okuduğunuz hafif zarar versiyonunda olduğu gibi). Ceza tazminatı 1,5 milyon dolar olarak belirlendi.*

General Asistance'a ceza belirleme sorununun bir anda çok daha kolaylaştığına dikkatinizi çekmek isteriz. Hatta aklınıza çoktan bir miktar gelmiş bile olabilir. Joan'a verilen ağır hasarla, diğer kızın maruz kaldığı hafif hasar arasındaki tezada karşılık gelen bir dolar tutarı çarpanı (ya da oranı) var. Dahası okuduğunuz tekil çapa (1,5 milyon dolar) dolar ceza ölçeğinin tamamını bir aralığa sabitlemeye yeterli. Şu ana dek değerlendirdiğiniz bu iki örnekten daha ağır ve daha hafif vakalara tazminat tayin etmek artık sizin için çok daha kolay.

Bir oran ölçeğinde yargılara varmak için çapalar gerekliyse, insanlara çapa sağlanmadığında ne olur? Stevens bu sorunun cevabını da raporladı. Deneyi yapan kişinin rehberliği olmadığında, insanlar ölçeği ilk kez kullandıklarında keyfî bir seçim yapmak zorunda kalıyorlar. O noktadan itibaren, yargılarını ilk cevaplarını bir çapa olarak kullanarak tutarlı bir şekilde veriyorlar.

Joan Clover vakasında tazminat belirlerken karşı karşıya bırakıldığınız görevin bir çapasız ölçeklendirme örneği olduğunu fark etmiş olabilirsiniz. Stevens'ın laboratuvarındaki çapasız gözlemciler

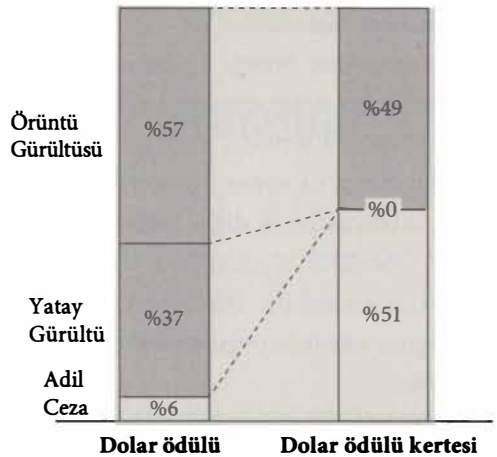
gibi siz de General Asistance'ın doğru cezası için keyfî bir karar verdiniz. Cezalandırıcı tazminat çalışmamızdaki katılımcılar da aynı sorunla karşı karşıya geldiler: Onlar da gördükleri ilk vaka hakkında keyfî bir ilk karar vermeye zorlandılar. Ancak sizin aksinize, onlar sadece bir keyfî karar vermekle kalmadı: Dokuz vaka için daha ceza tazminatları belirleyerek devam ettiler. Bu dokuz yargı keyfî değildi çünkü ilk çapalama kararıyla ve dolayısıyla birbirleriyle tutarlı hâle getirilebildiler.

Steven'ın laboratuvarının bulguları bireylerin ürettiği çapanın, daha sonraki dolar yargılarının mutlak değerleri üstünde büyük bir etkisi olması gerektiğini ama on vakanın göreceli pozisyonları üstünde hiçbir etkisi olmadığını ortaya koydu. Büyük bir ilk yargı, sonraki diğer bütün yargıların göreceli boyutlarını etkilemeden orantılı bir şekilde büyük olmasına neden oluyordu. Bu akıl yürütme şekli bizi şaşırtıcı bir sonuca ulaştırıyor: Umutsuzca gürültülü görünmelerine rağmen, dolar yargıları aslında yargıçların cezai niyetlerini yansıtır. Bu niyetleri keşfetmek için mutlak dolar değerlerinin yerine göreceli puanları koymamız yeterlidir.

Bu fikri test etmek için, gürültü analizini her dolar miktarını bir bireyin on yargısı arasındaki sırasıyla değiştirdikten sonra tekrarladık. En yüksek dolar tutarına 1, bir sonraki en yüksek tutara 2 dendi ve böyle devam etti. Dolar cezalarının sıralamaya dönüştürülmesi jüri üyesi düzeyindeki hataların tamamını ortadan kaldırır çünkü arada bir ortaya çıkan eşitlik dışında, sıralamanın 1'den 10'a dağılımı herkesi için aynıdır. (Merak ettiyseniz, anketin birden fazla versiyonu vardı çünkü her birey yirmi sekiz senaryodan on tanesi hakkında karar verdi. Analizi aynı on senaryoya cevap veren her katılımcı grubu için ayrı ayrı yürüttük ve bir ortalama raporladık.)

Sonuçlar çarpıcıydı: Yargılardaki gürültünün oranı %94'ten %49'a düştü (Şekil 14). Dolar miktarlarının sıralamalara dönüştürülmesi jüri üyelerinin faklı vakalarda uygun ceza konusunda

büyük ölçüde görüş birliği içinde olduklarını ortaya koydu. Hatta dolar olarak ceza sıralamaları, orijinal cezai niyet puanlamalarından biraz *daha az* gürültülüydü.



ŞEKİL 14: *Değer olarak gürültüye karşı sıralama olarak gürültü*

Talihsiz Bir Sonuç

Sonuçlar ana hatlarını çizdiğimiz teoriyle tutarlıdır. Bütün vakalar için dolar olarak ceza miktarı her jüri üyesinin gördükleri ilk vaka için seçtikleri keyfî sayıya çapalanmıştı. Vakaların göreceli sıralanması makul isabetlilik içeren tavırları yansıtır ve dolayısıyla çok gürültülü değildir ancak dolar cinsinden cezaların mutlak değerleri özde anlamsızdır çünkü ilk vakada seçilen keyfî sayılara bağlıdırlar.

İronik olarak, gerçek duruşmalarda jüri üyelerinin değerlendirdiği vaka gördükleri ilk ve tek vakadır. Amerikan yargı sistemi sivil jürilerin herhangi bir rehber çapadan faydalanmadan bir tutar tayin etmesini ister. Yasa jüri üyelerinin başka vakalardaki cezalandırıcı tazminatların boyutları hakkında bilgilendirilmesini alenen yasaklar. Yasada üstü örtülü olan varsayım, jüri üyeleri-

nin adalet duygusunun onları bir ceza üstünde kafa yormaktan doğru cezaya doğrudan yönlendireceği şeklindedir. Bu varsayım psikolojik saçmalıktır; insanlarda olmayan bir beceriyi varsayar. Adalet kurumları onlara işlerlik kazandıran insanların sınırlarının farkında olarak hareket etmelidir.

Cezalandırıcı tazminat örneği uç bir örnektir; profesyonel yargılar nadiren bu kadar umutsuz derecede muğlak ölçekler üstünde ifade edilir. Bununla birlikte, muğlak ölçekler yaygındır ve bu, cezalandırıcı tazminat çalışmasının içerdiği ve iş dünyasında, eğitimde, sporda, hükûmette ve diğer yerlerde uygulanabilir iki derse dikkat çeker. Öncelikle, ölçek seçimi yargıların gürültü miktarında büyük farklar yaratabilir. İkinci olarak, mutlak yargıların yerine göreceli yargılar koymak, uygulanabilir olduğu durumlarda, gürültüyü azaltabilir.

Ölçeklerden Bahsederken

"Yargılarımızda çok fazla gürültü var. Nedeni ölçeği farklı anlamamız olabilir mi?"

"O ölçek üstünde referans görevi üstlenecek bir çapa vaka üstünde anlaşabilir miyiz?"

BÖLÜM 16

Örüntüler

Bölüm 14'te üniversite not ortalamasını tahmin etmeye çalıştığınız, okumayı vaktinden çok önce söken Julie'yi hatırlıyor musunuz? Aşağıda daha detaylı bir tarifini bulacaksınız.

Julie tek çocuktu. Babası başarılı bir avukat, annesi ise mimardı. Julie yaklaşık üç yaşındayken babası onu evden çalışmaya mecbur bırakan bir otoimmün hastalığa yakalandı. Julie'yle çok fazla zaman geçirdi ve ona sabırla okumayı öğretti. Julie dört yaşında akıcı bir şekilde okuyabiliyordu. Babası ona aritmetik öğretmeyi de denedi ama bu Julie'ye zor geldi. Julie ilkokulda iyi bir öğrenciydi ama duygusal açıdan ilgiye muhtaç ve popüler olmaya çok uzaktı. Vaktinin büyük kısmını tek başına geçirdiği için en sevdiği amcasıyla kuş gözlemlerken aldığı bir ilhamla tutkulu bir kuş gözlemcisi oldu.

Annesi ve babası o on bir yaşındayken boşandılar ve boşanma Julie'yi derinden sarstı. Ders notları hızla düştü ve okulda sık sık patlamalar yaşamaya başladı. Lisedeyken biyoloji ve yaratıcı yazarlığın aralarında ol-

duğu bazı derslerde çok iyiydi. Fizikte çok başarılı olarak herkesi şaşırttı. Ama diğer derslerin çoğunu ihmal ettiği için liseden "İyi" dereceyle mezun oldu.

Başvurduğu prestijli okullara kabul edilmeyen Julie sonunda iyi bir devlet üniversitesine girdi ve orada çevre bilimleri üzerine lisans yaptı. Üniversitedeki ilk iki yılında sık sık duygusal karmaşalar yaşamaya devam etti ve düzenli olarak ot içti. Ancak dördüncü sömestirinde içinde güçlü bir tıp fakültesine gitme arzusu doğdu ve çalışmalarını çok daha fazla ciddiye almaya başladı.

Julie'nin diploma not ortalaması hakkında en iyi tahmininiz ne?

Problemler: Zor ve Kolay

Tahmin edileceği gibi bu problem (Julie 2.0 diyelim) çok daha zorlaştı. Julie 1.0 hakkında bütün bildiğiniz dört yaşındayken okumayı söktüğüydü. Tek bir ipucuyla, eşlemenin gücü görevini yaptı ve not ortalamasının sezgisel bir ortalaması hızla aklınızda belirdi.

Aynı genel istikameti işaret eden birden fazla ipucunuz olsaydı, eşleştirme yine kullanılabilirdi. Örneğin cazla uğraşan muhasebeci Bill'in tarifini okuduğunuzda edindiğiniz bilginin tamamı ("yaratıcı değil", "matematikte güçlü", "sosyal bilimlerde zayıf") tutarlı, basmakalıp bir resim çiziyordu. Benzer bir şekilde, Julie 2.0'ın hayatındaki olayların çoğu bir vaktinden önce gelişme ve üstün başarı hikâyesiyle tutarlı olsaydı (belki sadece "vasat" performansı işaret eden veri sayısı biraz daha az olsaydı) görev size bu kadar zor gelmezdi. Eldeki kanıt tutarlı bir resim çizdiği zaman, hızlı, Sistem 1 düşünmemiz resme anlam vermekte zorlanmaz. Bunun gibi basit yargı problemleri kolayca çözülür ve pek çok insan çözüm konusunda hemfikir olur.

Julie 2.0'da ise durum o kadar basit değil. Bu problemi zor yapan çok sayıda, çelişkili ipuçları. Beceri ve motivasyon göstergeleri var ama zayıf bir karakter ve vasat başarı işaretleri de söz konusu. Hikâye karmakarışık görünüyor. Ögeler tutarlı bir yoruma uydurulamadığı için kolayca anlamlandırılamıyor. Elbette tutarsızlık hikâyeyi daha az gerçekçi ve hatta mantıksız kılmıyor. Hayat genellikle hakkında anlatmaktan hoşlandığımız hikâyelerden daha karmaşıktır.

Çoklu, çelişkili ipuçları zor yargı sorunlarını tanımlayan muğlaklığı yaratır. Muğlaklık ayrıca karmaşık problemlerin neden basit olanlardan daha gürültülü olduğunu da açıklar. Kural basittir: Herhangi bir şeyi görmenin birden fazla yolu varsa, insanların onu görüş biçimleri farklılık gösterecektir. İnsanlar anlatılarının çekirdeğini oluşturmak için farklı kanıt parçaları seçebileceği için, çok sayıda olası sonuç vardır. Julie 2.0'ı anlamlandıran bir hikâye oluşturmakta zorlanıyorsanız başka okuyucuların sizinkilerden farklı yargıları haklı çıkaran farklı hikâyeler oluşturacaklarından emin olabilirsiniz. Örüntü gürültüsü yaratan farklılık budur.

Bir yargıdan ne zaman emin hissedersiniz? İki şartın karşılanması gerekir: İnandığınız hikâye kapsamlı olarak tutarlı olmalı ve başka cazip alternatif olmamalıdır. Kapsamlı tutarlılık, seçilen yorumun bütün detayları hikâyeye uyduğunda ve birbirini pekiştirdiğinde elde edilir. Elbette tutarlılığı, her ne kadar böylesi daha az zarifçe de olsa, uymayan şeyleri görmezden gelerek ve örtbas ederek de sağlayabilirsiniz. Alternatif yorumlarda da aynı durum geçerlidir. Bir yargı problemini "çözen" gerçek bir uzman sadece açıklayıcı hikâyesinin neden doğru olduğunu bilmekle kalmaz, diğer hikâyelerin neden hatalı olduğunu izah etmekte de bir o kadar akıcıdır. Yine, bir insan alternatifleri göz önünde bulundurmakta eksik kalarak ya da onları aktif olarak baskılayarak eşit güçte ama daha düşük kalitede güven edinebilir.

Güven konusuna bu bakışın ana iması, kişinin kendi yargısına öznel güveninin isabetliliği hiçbir şekilde garantilemediğidir. Dahası alternatif yorumların bastırılması (algıda çok incelenmiş bir süreç) *anlaşma yanılsaması* (bakınız Bölüm 2) adını verdiğimiz şeyi de doğurabilir. İnsanlar çıkarımlarına olası alternatifler hayal edemezlerse, doğal olarak diğer gözlemcilerin de aynı çıkarıma ulaşacağını varsayarlar. Elbette, bütün yargılarımızdan fazlasıyla emin olma talihine çok azımız sahibiz ve hepimiz belki de az önce Julie 2.0'ı okurken yaşadığınız gibi, bir belirsizlik tecrübesi yaşamışızdır. Her zaman yargılarımızdan çok emin değilizdir ama çoğu zaman olmamız gerekenden daha fazla emin oluruz.

Örüntü Gürültüsü: Sabit ya da Geçici

Örüntü hatasını bireyin, karar vericinin ayrı etkilerinin toplamıyla açıklanamayacağı bir vakayla ilgili yargısındaki bir hata olarak tanımlamıştık. Aşırı bir örnek olarak normalde merhametli olan bir yargıcın belli bir sanık türüne (örneğin trafik suçu işleyenlere) ceza verirken alışılmışın üstünde sert olması verilebilir. Ya da normalde temkinli bir yatırımcının heyecan verici yeni bir işin planlarını görünce, her zamanki temkinini bir kenara bırakması. Elbette çoğu örüntü hataları uç noktada değildir: Tekrarlı suç işleyen sanıklarla uğraşırken daha az merhametli davranan ya da genç kadınlar hakkında hüküm verirken her zamankinden daha merhametli olan yumuşak bir yargıçta da orta düzeyde bir örüntü hatası gözlemleriz.

Örüntü hataları geçici ve kalıcı faktörlerin bir kombinasyonundan doğar. Geçici faktörler, alakalı bir anda bir yargıcın ruh hâlinin iyi olması ya da yakın zamanda yaşanmış ve yargıcın hâlâ aklında olan talihsiz bir olay gibi durum gürültüsü kaynakları olarak tanımladığımız faktörleri içerir. Diğer faktörler daha kalıcıdır: Örneğin bir işverenin belli üniversitelerden mezun olan

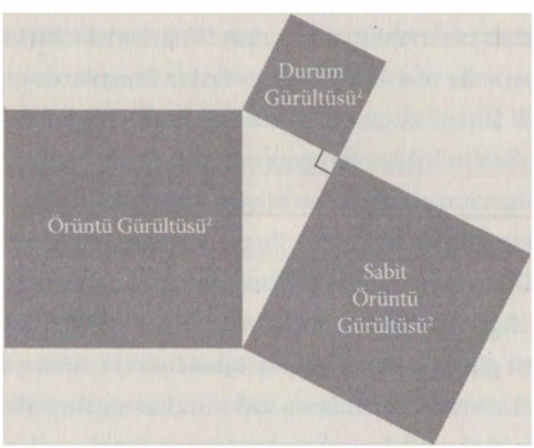
insanlara normalin üstünde bir yakınlık duyması ya da bir doktorun zatürre geçiren hastalara hastaneye yatmalarını önermeye normalden daha fazla meyilli olması gibi. Tekil bir yargıdaki bir hatayı basit bir denklemle tanımlayabiliriz:

Örüntü Hatası = Sabit Örüntü Hatası + Geçici (Durum) Hatası

Sabit örüntü hatası ve geçici (durum) hatası bağımsız oldukları ve aralarında bir bağıntı olmadığı için, yukarıdaki denklemi farklılıklarını analiz edecek şekilde genişletebiliriz.

(Örüntü Gürültüsü)^2 = (Sabit Örüntü Gürültüsü)^2 + (Durum Gürültüsü)^2

Hata ve gürültünün diğer ögelerinde yaptığımız gibi, bu denklemi grafik biçiminde, dik bir üçgenin kenarlarındaki karelerin toplamı olarak temsil edebiliriz (Şekil 15).



ŞEKİL 15: *Ayrıştırıcı örüntü gürültüsü*

Basit bir sabit örüntü gürültüsü vakası için, bir dizi puanlamaya dayanarak üst düzey yöneticilerin gelecekteki performanslarını öngören işe alım uzmanlarını düşünün. 9. Bölüm'de "yargıcın modeli"nden bahsetmiştik. Bireysel bir işe alım uzmanı her puanlamaya, kendi yargısındaki önemine karşılık gelen bir ağırlık verir. Ağırlıklar işe alım uzmanları arasında farklılık gösterir, liderlik bir uzman için daha önemliyken, bir diğeri için iletişim becerileri daha önemli olabilir. Bu tür farklılıklar işe alım uzmanlarının adayları sıralamasında farklılık üretir. Bu, sabit örüntü gürültüsü dediğimiz şeye bir örnektir.

Bireysel vakalara kişisel tepkiler ayrıca sabit ama fazlasıyla belirli olan örüntüler de üretebilir. Sizi Julie'nin hikâyesinin bazı yönlerine diğerlerinden daha fazla dikkat etmenize neyin yönlendirdiğini düşünün. Vakanın bazı detayları kendi hayat tecrübelerinizle örtüşebilir. Belki de Julie'de size, sürekli başarmanın kıyısına gelip sonunda başarısız olan ve sizin nedenini ergenlik yıllarından beri belirgin olan derin kişilik kusurlarına bağladığınız yakın bir akrabanızı anımsatan bir şey var. Tersine, Julie'nin hikâyesi sizde sorunlu bir ergenliğin ardından tıp fakültesine girmeyi başaran ve şimdi başarılı bir uzman olan yakın bir arkadaşınızla anılarınızı canlandırıyor da olabilir. Julie'nin farklı insanlarda uyandırdığı çağrışımlar kişiye özgü ve öngörülemezdir ama sabit olmaları olasıdır. Julie'nin hikâyesini geçen hafta okumuş olsaydınız, yine aynı insanları hatırlardınız ve hikâyesini aynı belirgin kişisel ışık altında görürdünüz.

Yargıların kalitesindeki kişisel farklılıklar örüntü gürültüsünün bir diğer kaynağıdır. Kimsenin (kendisinin bile) bilmediği kristal küre güçleri olan tekil bir tahminci düşünün. İsabetliliği onu pek çok vakada ortalama tahminden saptıracaktır. Sonuç verisinin yokluğunda bu sapmalar örüntü hataları olarak görülebilir. Yargılar doğrulanamaz olduğunda, üstün isabetlilik örüntü gürültüsüne benzer.

Örüntü gürültüsü ayrıca bir vakanın farklı boyutları hakkında geçerli yargılara varma becerisindeki sistematik farklılıklardan da doğar. Profesyonel spor takımları için seçim sürecini düşünün. Koçlar maçın farklı yönlerine odaklanırken, doktorlar sakatlanmaya yatkınlıkla, psikologlar ise motivasyon ve azimle ilgilenirler. Bu farklı uzmanlar aynı oyuncuları değerlendirirken, hatırı sayılır miktarda örüntü gürültüsü bekleyebiliriz. Benzer şekilde, daha genel beceri ve bilgilere sahip profesyoneller yargı görevinin bazı yönlerinde diğerlerinden daha yetenekli olabilirler. Bu tür durumlarda, örüntü gürültüsünün hatadan çok, insanların bildiklerindeki değişkenlik olarak tanımlanması daha doğru olur.

Profesyoneller tek başlarına karar verdiğinde, becerilerdeki değişkenlik basitçe gürültüdür. Ancak, yönetimin yargılara birlikte ulaşacak takımlar oluşturma imkânı olduğunda, becerilerin çeşitliliği bir potansiyel değere dönüşür çünkü farklı profesyoneller yargının farklı yönlerini ele alır ve birbirlerini tamamlarlar. Bu fırsatı ve yakalamak için gerekenleri Bölüm 21'de ele alacağız.

Daha önceki bölümlerde bir sigorta şirketinin müşterisinin ya da onu yargılayacak yargıca düşen sanığın iki piyangosundan bahsetmiştik. Bir meslektaşlar grubu içinden bir profesyoneli seçen ilk piyangonun profesyonelin yargılarının ortalama seviyesinden (seviye hatası) daha fazlasını seçtiğini artık biliyoruz. Piyango ayrıca bu profesyonele özgü değerlerin, tercihlerin, anıların, tecrübelerin ve çağrışımların kaleydoskop tarzında bir bileşimini de seçer. Bir yargıya vardığınız zaman, kendi yükünüzü de yanınızda taşırsınız. İşte oluşturduğunuz zihin alışkanlıklarını ya da akıl hocalarınızdan edindiğiniz bilgeliği de katarsınız. Öz güveninizi inşa eden başarıları ve tekrarlamamaya özen gösterdiğiniz hatalarınızı da yanınızda getirirsiniz. Ve beyninizin bir yerinde hatırladığınız, unuttuğunuz ve yok saymanın sorun olmadığını öğrendiğiniz resmî kurallar bulunur. Hiç kimse bütün bu açılardan tam olarak sizin gibi değildir; sabit örüntü hatalarınız size özeldir.

İkinci piyango ise yargıda bulunacağınız anı, ruh hâlinizi ve yargınızı etkilememesi gereken ama etkileyen dış şartları seçer. Bu piyango durum gürültüsü yaratır. Örneğin, Julie'nin vakasını okumadan kısa süre önce üniversite kampüslerinde uyuşturucu kullanımı hakkında bir gazete makalesi okuduğunuzu varsayın. Makale, hukuk fakültesine gitmeyi kafaya koymuş ve çok çalışan ama üniversitedeki ilk yıllarında uyuşturucu kullanırken açtığı gediği telafi edemeyen yetenekli bir öğrenciyi konu alıyordu. Bu hikâye zihninizde taze olduğu için, Julie'nin genel şansını değerlendirirken ot içme alışkanlığı daha fazla dikkatinizi çeker. Öte yandan, belki de Julie'nin sorusuyla birkaç hafta sonra karşılaşsanız, makaleyi hatırlamayacaktınız bile (ve elbette, vakayı dün okumuş olsaydınız makaleden haberiniz bile olmayacaktı). Gazete makalesi okumanın etkisi geçicidir; bu, durum gürültüsüdür.

Bu örneğin gösterdiği gibi sabit örüntü gürültüsü ve durum gürültüsü dediğiniz sabit olmayan değişken arasında keskin bir süreksizlik yoktur. Ana fark bir kişinin vakanın bazı yönlerine kendine has duyarlılığının kalıcı ya da geçici olmasıdır. Örüntü gürültüsünün tetikleyicileri kişisel tecrübe ve değerlerimize dayandığı zaman, örüntünün sabit olmasını, benzersizliğimizin bir yansıması olmasını bekleyebiliriz.

Kişilik Benzetmesi

Belli insanların, belirli özellik ve özellik kombinasyonlarına tepkilerinin benzersizliği sadece sezgisel değildir. Bunu anlamak için, hepimizin çok iyi bildiği bir başka karmaşık özellik kombinasyonunu düşünebiliriz: etrafımızdaki insanların kişilikleri. Aslında, bir yargıcın bir vaka hakkında bir yargıya varması daha geniş bir alan olan kişilik araştırması alanındaki özel bir vaka olarak görülmelidir: Bir durumda insan nasıl davranır? Yargı hakkında

konunun onlarca yıla yayılan daha geniş kapsamlı ve yoğun incelenmesinden öğrenilecek şeyler hâlâ var.

Psikologlar çok uzun zamandır kişilikteki bireysel farkları anlamak ve ölçmek için çabalıyorlar. İnsanlar birbirlerinden çok çeşitli şekillerde farklılık gösteriyorlar: Çok önce yapılan, sözlükte kişiliği tarif edebilecek terimler taramasının sonucunda sekiz bin sözcük saptanmıştı. Bugün baskın kişilik modeli, Beş Büyük Model'i, Beş Büyük'ten her birinin ayırt edilebilir özellikler aralığını kapsadığı beş grupta (dışa dönüklük, uyumluluk, sorumluluk, tecrübeye açıklık ve nörotizm) topluyor. Bir kişilik özelliği gerçek davranışların kestiricisi olarak anlaşılıyor. Kişi sorumlu olarak tanımlandığında, onda örtüşen davranışlar (dakik olmak, sözlerini yerine getirmek vs.) gözlemlemeyi bekleriz. Ve eğer Andrew saldırganlık ölçümünde Brad'den daha yüksek puan aldıysa, çoğu durumda Andrew'un Brad'den daha saldırgan davrandığını gözlemlemeliyiz. Bununla birlikte, aslında, belirli davranışları öngörmede geniş özelliklerin geçerliliği hayli sınırlıdır: 0,30luk bir korelasyon (YU = %60) yüksek kabul edilir.

Sağduyu davranışın kişilik tarafından yönlendirilmekle birlikte *durumlardan* da etkilendiğini düşünmeyi gerektirir. Bazı durumlarda hiç kimse saldırgan değildir, diğerlerinde ise herkes saldırgandır. Bir yakınını kaybetmiş bir arkadaşlarını teselli ederken ne Andrew ne de Brad saldırgan davranır. Ancak bir futbol karşılaşmasında ikisi de bir miktar saldırganlık sergiler. Kısaca –ve şaşırtıcı olmayan bir şekilde– davranışlar kişiliklerin *ve* durumların bir fonksiyonudur.

İnsanları eşsiz ve sonsuzca ilginç kılan şey, kişilik ve durumun buluşmasının mekanik, ek bir işlev olmamasıdır. Örneğin, üç aşağı beş yukarı saldırganlığı tetikleyen durumlar herkes için aynı değildir. Andrew ve Brad ortalamada eşit derecede saldırgan olsalar bile, her bağlamda illa eşit saldırganlık sergilemezler. Belki Andrew yaşıtlarına karşı saldırgan ama üstlerine karşı daha mü-

layimken Brad'in saldırganlık seviyesi hiyerarşik seviyeye duyarlı değildir. Belki Brad eleştirildiği zaman saldırganlaşmaya daha yatkın ve fiziksel olarak tehdit altında hissettiğinde normalden daha kontrollüdür.

Bu durumlara karşı kişiye özel tepki örüntüleri, zaman içinde sabit ve dengede kalmaya yatkındır. Her ne kadar geniş bir özellikle tanımlamaya katkı sağlamasalar da bir insanın *kişiliği* olarak adlandırdığımız şeyin büyük kısmını oluştururlar. Andrew ve Brad bir saldırganlık testinde aynı puanı paylaşıyor olabilirler ama saldırganlık tetikleyicilerine ve bağlamlara tepki örüntülerinde benzersizdirler. Bir özellik seviyesini paylaşan –örneğin eşit seviyede dikbaşlı ya da eşit derecede cömert– iki insan, aynı ortalamaya sahip olan ama farklı durumlara karşı illa aynı tepki örüntülerine sahip olmayan iki davranış dağılımıyla tanımlanmalıdır.

Bu kişilik tartışması ve sunduğumuz yargı modeli arasındaki paralelliği artık görüyor olmalısınız. Yargıçlar arasındaki seviye farklılıkları, birden fazla durumdaki davranışlar ortalamasını temsil eden kişilik özellikleri puanlarındaki farklılıklara karşılık gelir. Vakalar durumlara benzer. Bir kişinin belli bir sorunla ilgili yargısı o kişinin ortalama seviyesinden sadece bir miktar öngörülebilir, tıpkı belirli davranışların kişilik özelliklerinden sadece belli bir miktarda öngörülebilmesi gibi. Bireylerin yargılarına göre sıralaması bir vakadan diğerine büyük ölçüde farklılık gösterir çünkü insanlar her vakada buldukları özelliklere ve özellik kombinasyonlarına tepkilerinde farklılık sergilerler. Yargılayan karşı benzersiz bir duyarlılık örüntüsü olmanın yanı sıra buna bağlı olarak vakaların yargılanmasında da benzersiz bir örüntüdür.

Kişiliğin benzersizliği normalde kutlanacak bir şeydir ama bu kitap farklılığın sorun yaratabildiği ve gürültünün hata olduğu profesyonel yargıları konu alıyor. Benzetmenin maksadı, her ne kadar açıklamak konusunda çok umutlu olmasak ve diğerlerinden

belirgin şekilde faklı kararlar alan bireyler bunu izah edemeseler de yargıda örüntü gürültüsünün rastgele olmadığını göstermektir.

Örüntü Gürültüsünden Bahsederken

"Çıkarımından emin görünüyorsun ama bu kolay bir sorun değil: Çok farklı yönleri işaret eden ipuçları var. Kanıtın alternatif yorumlarını gözden kaçırmış olabilir misin?"

"Sen ve ben aynı adayla mülakat yaptık ve normalde ikimiz eşit derecede zorlu mülakatçılarız. Ama ikimiz bambaşka yargılara ulaştık. Bu örüntü gürültüsü nereden kaynaklanıyor?"

"İnsanların kişiliklerinin benzersizliği onlara yenilik ve yaratıcılık becerisi verir ve yanlarında olmayı ilginç ve heyecan verici kılar. Konu karar vermeye geldiğinde ise, bu benzersizlik bir artı değer olmaz."

BÖLÜM 17

Gürültü Kaynakları

Artık yargının olduğu her yerde gürültünün de olduğu konusunda hemfikir olduğunuzu umuyoruz. Ayrıca sizin için artık düşündüğünüzden daha fazlasının olmadığını da umuyoruz. Gürültüyle ilgili bu düstur projeye başlarken bizi motive etti ama konu hakkında düşüncemiz üstünde çalıştığımız yıllar içinde ilerleme kaydetti. Şimdi gürültünün ögeleri, bu ögelerin gürültünün genel resmindeki göreceli önemleri ve yargının incelenmesinde gürültünün yeri hakkında aldığımız başlıca dersleri gözden geçireceğiz.

Gürültünün Ögeleri

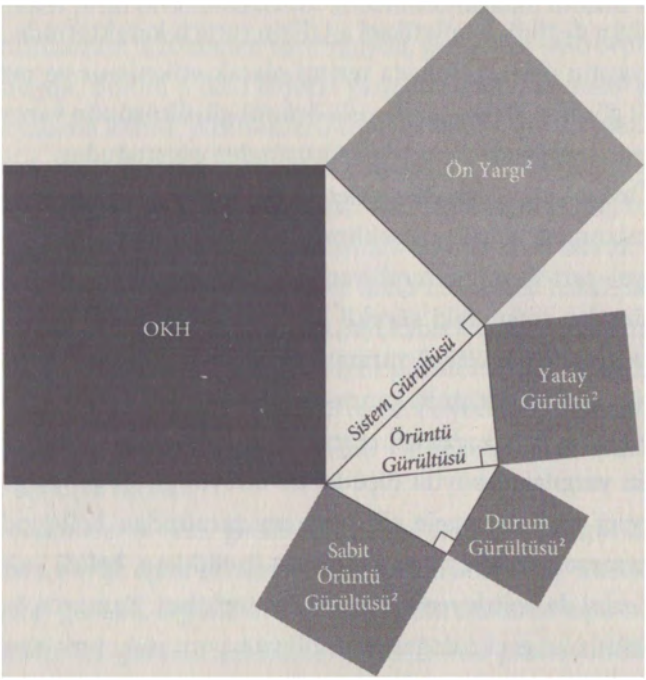
Şekil 16'da, Bölüm 5, 6 ve 16'da bahsettiğimiz üç denklemin kombine grafik sunumlarını bulabilirsiniz. Bu şekil, hatanın peş peşe üç çözümlemesini resmediyor.

hatanın ön yargısı ve sistem gürültüsü

sistem gürültüsünün yatay gürültüsü ve örüntü gürültüsü

örüntü gürültüsünün sabit örüntü gürültüsü ve durum gürültüsü

Şimdi OKH'nin ön yargı karelerine ve gürültünün ele aldığımız üç ögesine nasıl ayrıldığını görebilirsiniz.



ŞEKİL 16: *Hata, ön yargı ve gürültünün ögeleri*

Araştırmamıza başladığımızda ön yargının ve gürültünün toplam hatadaki göreceli ağırlıklarına odaklanmıştık. Çok geçmeden gürültünün sıklıkla hatanın ön yargıdan daha büyük bir ögesini oluşturduğu ve daha detaylı olarak incelenmeyi hak ettiği sonucuna ulaştık.

Başlangıçta gürültünün ögeleriyle ilgili düşünme şeklimize çok sayıda insanın çok sayıda vakada bireysel yargılara vardığı

karmaşık gürültü denetimlerinin yapısı rehberlik ediyordu. Federal yargıçlarla ilgili çalışma ve cezalandırıcı tazminatla ilgili inceleme birer örnekti. Bu çalışmaların verileri yatay gürültü için sağlam tahminler sağladı. Öte yandan her katılımcı her vakada yargıya vardığı ama bunu sadece bir kez yaptığı için örüntü hatası adını verdiğimiz artık hatanın geçici mi yoksa sabit mi olduğunu bilmek mümkün değildi. İstatistiksel analizin tutucu karakterinde, artık hata yaygın olarak bir hata terimi olarak etiketlenir ve rastgele olarak görülür. Başka bir deyişle örüntü gürültüsünün varsayılan yorumu, tamamen durum gürültüsünden oluştuğudur.

Örüntü gürültüsünün geleneksel olarak rastgele hata olarak yorumlanması, düşünme şeklimizi uzun süre kısıtladı. Yatay gürültüye –sert ve merhametli yargıçlar ya da iyimser ve kötümser tahminciler arasındaki sürekli farka– odaklanmak doğal geldi. Ayrıca durum gürültüsü yaratan alakasız ve geçici şartların yargılar üstündeki etkisinin kanıtı da ilgi çekiciydi.

Kanıtlar bizi kademeli olarak farklı insanların yarattığı gürültülü yargıların büyük ölçüde, ne bireyin genel ön yargısı ne de geçici ya da rastgele olan bir şey tarafından belirlendiğini fark etmeye itti: bireylerin belli bazı özelliklere, belirli vakalara tepkilerini de belirleyen ısrarlı kişisel tepkileri. Zamanla örüntü gürültüsünün geçici doğasıyla ilgili varsayımımızı terk etmemiz gerektiği sonucuna vardık.

Kısıtlı bir örnekler grubundan aşırı genelleme yapmamaya dikkat etmek istesek de bir araya topladığımız çalışmalar, sabit örüntü gürültüsünün sistem gürültüsünün diğer ögelerinden daha önemli olduğunu ortaya koyuyor. Aynı çalışmada hata ögelerinin tam bir resmini nadiren bulduğumuz için, bu çekimser sonucu oluşturmak bir miktar üçgenleme gerektiriyor. Kısaca, bildiklerimiz ve bilmediklerimiz şunlar:

Ögeleri Boyutlandırmak

Önce, yatay gürültü ve örüntü gürültüsünün göreceli ağırlıkları konusunda çok sayıda tahminimiz var. Toplamda, örüntü gürültüsünün payı yatay gürültüden daha büyük görünüyor. Örneğin, Bölüm 2'deki sigorta şirketinde, risk değerlendirme uzmanlarının belirledikleri prim ortalamalarının arasındaki farklar, toplam sistem gürültüsünün %20'sinden sorumluydu, geri kalan %80 örüntü gürültüsüydü. Bölüm 6'daki federal yargıçlar arasında yatay gürültü (ortalama katılık farklılıkları) toplam sistem gürültüsünün yarısından biraz daha azını temsil ediyordu; örüntü gürültüsü daha büyük bir ögeydi. Cezalandırıcı tazminat deneyinde, toplam sistem gürültüsü miktarı kullanılan ölçeğe göre büyük ölçüde değişiyordu (cezai niyet, öfke ve dolar cinsinden hasar) ama örüntü gürültüsünün bu toplamdaki payı kabaca sabitti: Çalışmada kullanılan üç ölçek için toplam sistem gürültüsünün %63, %62 ve %61'inden sorumluydu. Kısım 5'te inceleyeceğimiz özellikle kişisel kararlarla ilgili diğer çalışmalar da henüz teyide muhtaç olan bu sonuçla tutarlı.

Bu çalışmalarda yatay gürültünün genel olarak sistem gürültüsünün daha büyük ögesi olmaması başlı başına önemli bir mesajdır, çünkü yatay gürültü, örgütlerin (bazen) gürültü denetimi yapmadan izleyebildiği tek gürültü şeklidir. Vakalar farklı profesyonellere rastgele dağıtıldığında, kararlarının ortalama seviyesi arasındaki farklar sistem gürültüsüne kanıt oluşturmaktadır. Örneğin patent bürolarıyla ilgili çalışmada, müfettişlerin ortalama patent tanıma eğilimlerinde, akabinde bu patentler hakkında yaşanan ihtilaflar üstünde etkisi olan büyük farklılıklar gözlemlendi. Benzer şekilde, çocuk koruma hizmetlerindeki vaka memurları da çocukları devlet bakımına yerleştirme eğilimleri konusunda, çocuğun uzun vadeli esenliği üstünde sonuçlar yaratan farklılıklar sergilerler. Bu gözlemler sadece yatay gürültü tahminine dayalıdır. Örüntü gürültüsü yatay gürültüden çoksa zaten şoke edici olan bu bulgular gürültü

sorununu en az iki faktörle olduğundan daha az gösterir. (Teyide muhtaç olan bu kuralın istisnaları var. Sığınma talebi kararlarında yargıçların arasındaki skandal düzeydeki farklılık, neredeyse kesin olarak, örüntü gürültüsünden çok yatay gürültüye bağlı ve şüphemiz örüntü gürültüsünün de büyük olduğu yönünde.)

Bir sonraki adım, örüntü gürültüsünü iki ögesine ayırarak analiz etmektir. Durum gürültüsünden çok, sabit örüntü gürültüsünün baskın öge olduğunu varsaymak için birçok nedenimiz var. Federal yargıçların hükümlerinin denetimi bizim akıl yürütme şeklimizi resmediyor. İşe bütün örüntü gürültüsünün geçici olduğu olasılığıyla başlayalım. Bu varsayımla, hüküm vermek zaman içinde bize mantıksız gelen bir düzeyde dengesiz ve tutarsız olacaktır: *Aynı yargıcın aynı vakada* farklı durumlarda vereceği hükümler arasındaki ortalama farkın yaklaşık 2,8 yıl seviyesinde olmasını beklememiz gerekir. Yargıçlar arasında ortalama hükmün değişkenliği zaten şoke edicidir. Belli bir yargıcın farklı durumlardaki hükümlerde aynı çeşitliliği sergilemesi korkunç olur. Yargıçların farklı sanıklara ve farklı suçlara tepkilerinde farklılık gösterdikleri ve bu farkların son derece kişisel ama sabit olduğu sonucunu çıkarmak daha akla yatkın görünüyor.

Örüntü gürültüsünün ne kadarının sabit ne kadarının durum gürültüsü olduğunu daha kesin saptamak için, aynı yargıcın her vaka için iki bağımsız değerlendirme yaptığı çalışmalara ihtiyacımız var. Belirttiğimiz gibi, yargı incelemelerinde iki bağımsız yargı elde etmek genel olarak imkânsızdır, çünkü bir vakanın ikinci yargısının ilkinden gerçek anlamda bağımsız olmasını garantilemek zordur. Özellikle yargı karmaşık olduğunda, bireyin bir sorunu tanıyıp ilk yargısını tekrarlama olasılığı çok yüksektir.

Princeton'da Alexander Todorov'un liderlik ettiği bir araştırmacı grubu bu sorunu aşmak için akıllı deney teknikleri tasarladı. Bireylerin anket doldurmak gibi kısa süreli hizmetler sağladığı ve harcadıkları zamana karşılık ücret aldıkları Amazon Mechanical

Türk sitesinden katılımcılar kiraladılar. Bir deneyde katılımcılar yüz resimleri (bir bilgisayar programı tarafından üretilmiş ama gerçek insan yüzlerinden kesinlikle ayırt edilemeyen resimler) gördüler ve bu yüzleri sevimlilik ve güvenilirlik gibi çeşitli özelliklere göre puanladılar. Deney bir hafta sonra aynı yüzlerle ve aynı katılımcılarla tekrarlandı.

Bu deneyde yargıçların hükümleri gibi profesyonel yargılardan daha az konsensüs beklemek normaldir. Bazı insanların son derece çekici, diğerlerinin son derece itici olduğu konusunda herkes hemfikir olabilir ama geniş bir aralık üzerinde, yüzlere verilen tepkilerin büyük ölçüde kişiye özel olmasını bekleriz. Gerçekten de gözlemciler arasında az miktarda anlaşma sağlandı: Örneğin güvenilirlik puanlamasında, resimler arasındaki farklar yargıların değişkenliğinin sadece %18'inden sorumluydu. Farklığın kalan %82'si gürültüydü.

Bu yargılarda daha az istikrar beklemek de doğrudur, çünkü soruları çevrim içi yanıtlamak için para alan katılımcıların vardığı yargıların kalitesi genellikle profesyonel ortamlardakilerden önemli ölçüde düşüktür. Bununla birlikte, gürültünün en büyük ögesi sabit örüntü gürültüsüydü. İkinci en büyük ögesi ise yatay gürültüydü, yani gözlemcilerin güvenilirlik için verdiği puanların ortalamaları arasındaki farklardı. Hâlâ önemli olmakla birlikte, durum gürültüsü en küçük ögeydi.

Araştırmacılar katılımcılardan arabalar ve yiyecekler arasından tercihleri ya da profesyonel yargı dediğimiz şeye daha yakın soruları cevaplamak gibi başka yargılarda bulunmalarını istedikleri zaman yine aynı sonuçlara ulaştılar. Örneğin, Bölüm 15'te ele aldığımız cezalandırıcı tazminat çalışmasının bir kopyasında, katılımcılar bir hafta arayla iki farklı seferde, on kişisel yaralanma vakasında cezai niyetlerini puanladılar. Burada yine, sabit örüntü gürültüsü en büyük ögeydi. Bütün bu çalışmalarda bireyler genel olarak birbirleriyle aynı fikirde değildiler ama yargılarında istikrarlarını

korudular. Araştırmacıların kelimeleriyle "konsensüssüz tutarlılık" sabit örüntü gürültüsü için çok açık kanıtlar sağlar.

Sabit örüntülerin rolü için en güçlü kanıt, Bölüm 10'da bahsettiğimiz kapsamlı kefalet yargıçları çalışmasından geliyor. Bu sıra dışı çalışmanın bir kısmında, yazarlar her yargıcın kefalet hakkı tanımak ve tanımamak arasında karar verirken, eldeki ipuçlarını nasıl kullandığını taklit eden bir istatistik modeli yarattılar. 173 yargıcın özel yapım modellerini ürettiler. Sonra, simüle edilen yargıçlara 141.833 davanın her biri için 173'er adet, yani toplamda 24 milyondan fazla karar aldırdılar. Yazarlar talebimiz üzerine cömertlik göstererek, birbirinden farklı kararları üç ögeye ayırdıkları özel bir analiz gerçekleştirdiler: her dava için ortalama kararların "gerçek" çeşitliliği, yargıçlar arasında kefalet hakkı tanımaya yatkınlıkları açısından farklılıklarının yarattığı yatay gürültü ve geri kalan örüntü gürültüsü.

Bu analizin bizim argümanımızla ilgisi, örüntü gürültüsünün bu çalışmada ölçüldüğü hâliyle tamamen sabit olması yüzündendir. Durum gürültüsünün rastgele değişkenliği temsil edilmedi çünkü bu bir yargıcın kararını öngören *modellerin* bir analiziydi. Sadece doğrulanabilecek kadar sabit bireysel tahmin kuralları dâhil edildi.

Sonuç tartışmaya yer bırakmayacak kadar açıktı: Bu sabit örüntü gürültüsü, yatay gürültüden dört kat büyüktü (toplam farklılığın %26'sını sabit örüntü gürültüsü, %7'sini ise yatay gürültü açıklıyordu). Saptanabilen sabit ve kişiye has bireysel yargı örüntüleri, araştırmanın tamamındaki katılık farklılıklarından çok daha büyüktü.

Bütün bu kanıtlar Bölüm 7'de tartıştığımız durum gürültüsü araştırmasıyla tutarlı. Durum gürültüsünün varlığı şaşırtıcı ve hatta rahatsız edici bile olsa, kişinin kendi içindeki değişkenliklerin kişiler arası farklılıklardan daha büyük olduğuna dair bir gösterge yok. Sistem gürültüsünün en önemli ögesi başlangıçta

yok saydığımız ögedir: sabit örüntü ögesi, yargıçlar arasında belli vakalardaki yargılarında bulunan değişkenlik.

İlgili araştırmaların göreceli azlığı göz önüne alındığında, çıkarımlarımız, gürültü konusunu ele almamız ve üstesinden gelmemiz konusundaki düşüncelerimizde bir değişikliği yansıtmaktalar. En azından prensipte, yatay gürültü ya da yargıçlar arasındaki yaygın farklılıklar ölçülmesi ve ele alınması nispeten kolay bir sorun olmalı. Anormal derecede sert puanlayıcılar, "temkinli" çocuk gözetim memurları, "risk almaktan kaçınan" kredi memurları varsa onları çalıştıran örgütler yargılarının ortalama seviyesini eşitlemeyi hedef alabilir. Örneğin üniversiteler bu sorunu hocalarından her sınıf içinde önceden belirlenmiş bir not dağılımına sadık kalmalarını talep ederek ele alabilir.

Ne yazık ki artık fark ettiğimiz üzere, yatay gürültüye odaklanmak bireysel farklılıklarla ilgili çok büyük bir şeyi gözden kaçırmaya neden oluyor. Gürültü çoğunlukla seviye farklarının değil, etkileşimlerin ürünüdür. Farklı yargıçların belli sanıkları, farklı öğretmenlerin belli öğrencileri, farklı sosyal hizmet görevlilerinin belli aileleri, farklı liderlerin geleceğe dair belli vizyonları nasıl ele aldığı gibi. Gürültü benzersizliğimizin, "yargılama kişiliğimizin" bir yan ürünüdür. Yatay gürültüyü azaltmak yine de değerli bir amaçtır ama sadece bu hedefe ulaşmak sistem gürültüsü sorununun büyük kısmını çözümsüz bırakır.

Hatayı Açıklamak

Gürültü hakkında söyleyecek çok şey bulduk ama konu kamu farkındalığının ve yargı ve hata tartışmalarının tamamen dışında kalıyor. Varlığının ve onu üreten çok sayıda mekanizmanın kanıtlarına rağmen, gürültü yargıda ancak nadiren önemli bir faktör olarak görülür. Bu nasıl mümkün olabilir? Neden ön yargıyı rutin olarak suçlarken, kötü yargıların izahında gürültü hiç aklımıza

gelmez? Yaygınlığına rağmen, gürültüye bir hata kaynağı olarak kafa yormak neden bu kadar sıra dışıdır?

Bu bulmacanın kilidi, hataların ortalaması (ön yargı) ve hataların değişkenliğinin (gürültü) hata denkleminde eşit rol oynamasına rağmen, onları farklı şekillerde düşünmemizdir. Ve etrafımızdaki dünyaya normal anlam verme şeklimiz gürültünün rolünü tanımamızı imkânsız kılar.

Bu kitabın daha önceki kısımlarında her ne kadar olayları gerçekleşmelerinden önce öngörmesek de geriye dönüp baktığımızda kolayca anlamlandırabildiğimizden bahsetmiştik. Normallik vadisinde olaylar şaşırtıcı değildir ve kolaylıkla açıklanabilirler.

Aynı şey yargılar için de söylenebilir. Diğer olaylar gibi yargılar ve kararlar da çoğunlukla normallik vadisinde gerçekleşir; genellikle bizi şaşırtmazlar. Öncelikle, tatmin edici sonuçlar üreten yargılar normaldir ve nadiren sorgulanırlar. Frikik için seçilen futbolcu gol attığı zaman, kalp ameliyatı başarıyla sonuçlandığı zaman ya da yeni kurulan bir iş iyi gittiği zaman, karar alıcıların tercihlerini dayandırdıkları nedenlerinin doğru olduğunu varsayarız. Sonuçta haklı çıkmışlardır. Şaşırtıcı olmayan her hikâye gibi, bir başarı hikâyesi de kendini sonucu bilindiği zaman açıklar.

Öte yandan anormal sonuçları –kötü olanları ve zaman zaman karşılığını veren şaşırtıcı iş kumar gibi şaşırtıcı derecede iyi olanları– açıklama ihtiyacı da duyarız. Hataya ya da özel bir yeteneğe dayandırılan açıklamalar hak ettiklerinden çok daha popülerdir çünkü geçmişin önemli kumarları, sonuçları bildiğinde, kolayca dâhice eylemlere ya da çılgınlıklara dönüşebilir. *Temel yükleme hatası* denilen çok yaygın bir psikolojik ön yargı, şans ya da objektif şartlarla daha iyi açıklanan eylem ve sonuçlar için ajanlara suç ya da övgü atfetme eğilimidir. Bir başka ön yargı, geri görüş ön yargısı, yargıları, beklenemeyecek sonuçların geriye dönüp bakıldığında kolayca öngörülebilir görünmesini sağlayacak şekilde çarpıtır.

Yargı hataları için açıklamalara sık rastlanır; yargılar için nedenler bulmak, olaylara nedenler bulmaktan daha kolaydır. Her zaman kararları alan insanların gerekçelerini yardıma çağırabiliriz. Bu yetmezse, beceriksizliklerini suçlayabiliriz. Kötü kararların son yıllarda çok rağbet gören bir başka açıklaması da psikolojik ön yargıdır.

Psikoloji ve davranış ekonomisi alanlarında yapılan önemli bir araştırma, psikolojik ön yargılar için uzun bir liste oluşturdu: planlama yanılgısı, aşırı güven, kayıptan kaçınma, bağış etkisi, statüko eğilimi, geleceğin aşırı aza indirgenmesi ("şimdiki zaman ön yargısı") ve elbette çeşitli insan kategorilerinin lehine ya da aleyhine ön yargıları da içeren daha pek çokları. Bu ön yargıların her birinin yargı ve kararları etkilemesinin olası olduğu şartlar hakkında elimizde birçok bilgi var ve bir karar alma gözlemcisinin ön yargılı düşünmeyi gerçek zamanda ayırt etmesine olanak sağlayacak bilgiye de sahibiz.

Psikolojik ön yargı, ön yargı önceden tahmin edilebilmiş ya da gerçek zamanda saptanabilmiş olsa, bir yargı hatasına makul bir neden açıklaması olabilir. Olgudan sonra saptanan bir psikolojik ön yargı ise, gelecek hakkında da bir öngörü sağlaması durumunda, çekimser olmakla birlikte, faydalı bir açıklama sağlayabilir. Örneğin, güçlü bir kadın adayın bir pozisyon için reddedilmesi, aynı komitenin gelecekteki atamalarının onaylayacağı ya da çürüteceği daha genel bir cinsiyet yanlılığı hipotezini düşündürebilir. Tam tersine, sadece bir olaya uyan bir nedensel açıklama düşünün: "Başarısız oldukları o örnekte, aşırı öz güvenli davranmış olmalılar." Bu beyan tamamen mantıksız olmakla birlikte, hayli tatmin edici olabilecek bir anlayış yanılsaması sağlar. İşletme fakültesi hocası Phil Rosenzweig ön yargılar açısından boş açıklamaların işte alınan sonuçların tartışılmasında çok yaygın olduğunu ikna edici bir şekilde savunmuştur. Bu boş açıklamaların popülerliği

tecrübeyi anlamlandıracak nedensel hikâyelere yaygın bir ihtiyaç olduğunu doğrulamaktadır.

Gürültü İstatistikidir

Bölüm 12'de belirttiğimiz gibi, normal düşünme şeklimiz nedenseldir. Doğal olarak belirli bir şeye dikkat eder, bireysel vakalar hakkında başarısızlıkların sıklıkla hatalara, hataların da ön yargılara dayandırıldığı nedensel açıdan tutarlı hikâyeleri takip eder ve yaratırız. Kötü kararların açıklanmasındaki bu kolaylık hata açıklamalarımızda gürültüye hiç yer bırakmaz.

Gürültünün görünmezliği nedensel düşünmenin direkt bir sonucudur. Gürültü doğası gereği istatistikidir: Sadece benzer yargılar bütününü istatistiksel olarak düşündüğümüz zaman görünürlük kazanır. O zaman ise gözden kaçırılması çok zorlaşır: Ceza kararları ve sigorta primleriyle ilgili geriye dönük istatistiklerdeki değişkenliktir. Siz ve başkaları gelecekteki bir sonucu değerlendirirken ortaya çıkan olasılık yelpazesidir. İsabetli atışların hedef tahtasındaki dağılma şeklidir. Gürültü nedensel olarak hiçbir yerde, istatistiksel olarak ise her yerdedir.

Ne yazık ki istatistiksel bakış açısını benimsemek kolay değildir. Gözlemlediğimiz olaylar için hiç zorlanmadan nedenler üretiriz ama onları istatistiksel olarak düşünmek öğrenilmesi gereken bir şeydir ve çaba gerektirir. Nedenler doğal, istatistikler ise zordur.

Sonuç, ön yargı ve gürültüyü hata kaynakları olarak görmemizde dikkat çekici bir dengesizliktir. Giriş düzeyinde psikoloji biliyorsanız, muhtemelen çarpıcı ve bol detaylı bir figürün belirsiz bir arka planda öne çıktığı çizimleri hatırlarsınız. Arka planın önünde küçük kaldığı zaman bile, dikkatimiz o figüre sabitlenir. Figür/zemin sunumları ön yargı ve gürültüyle ilgili sezgilerimiz için çok uygun bir mecazdır: Ön yargı dikkat çekici figürdür,

gürültü ise dikkat etmediğimiz arka plan. Yargımızdaki büyük bir kusurdan büyük ölçüde bihaber kalmamız bu yüzdendir.

Gürültü Kaynaklarından Bahsederken

"Ortalama seviye yargılardaki farkları kolayca görürüz ama görmediğimiz örüntü gürültüsü ne kadar büyüktür?"

"Bu yargıya ön yargıların neden olduğunu söylüyorsun ama sonuç farklı olsaydı yine aynı şeyi söyler miydin? Ve gürültü olup olmadığını anlayabiliyor musun?"

"Olması gerektiği gibi, ön yargıları azaltmaya odaklıyız. Gürültüyü azaltmak konusunda da endişelenmeliyiz."

KISIM 5

Yargıları İyileştirmek

Bir örgüt profesyonellerinin aldığı kararları nasıl iyileştirebilir? Özellikle bir örgüt, yargı gürültüsünü nasıl azaltabilir? Bu soruları cevaplamak sizin sorumluluğunuz olsaydı nasıl bir yol izlerdiniz?

Gerekli bir ilk adım örgütün profesyonel yargılardaki gürültünün dikkati hak eden bir mesele olduğunu kabul etmesini sağlamaktır. Bu noktaya ulaşmak için bir gürültü denetimi (detaylı bir tarif için bakınız Ek A) gerekir. Gürültü denetimlerinde, çok sayıda yargıç aynı sorunları karara bağlar. Gürültü bu yargılardaki farklılıktır. Bu farklılığın beceriksizliğe dayandırılabileceği örnekler olacaktır: Bazı yargıçlar neden bahsettiklerini bilir, diğerleri bilmez. Arada böyle bir beceri farkı olduğunda (ya genel olarak ya da belli vaka türlerinde) öncelik elbette bu eksik becerileri iyileştirmek olmalıdır. Ama gördüğümüz gibi, becerikli ve iyi eğitimli profesyonellerin yargılarında bile büyük miktarda gürültü olabilmektedir.

Sistem gürültüsünün miktarı ele alınmaya değerse, yargının yerine kurallar ya da algoritmalar koymak, gürültüyü tamamen bertaraf edeceği için, düşünmeniz gereken bir seçenektir. Ama kuralların da kendi sorunları vardır (Kısım 6'da göreceğimiz gibi)

ve YZ'nin en hevesli destekçileri bile algoritmaların insan yargısının evrensel bir alternatifi olmadığı ve yakın zamanda olmayacağı konusunda hemfikirdir. Yargıları iyileştirme görevi her zamanki kadar acildir ve kitabın bu kısmında bu konuyu ele alacağız.

Yargıları iyileştirmenin akla yatkın bir yolu elbette mümkün olan en iyi insan yargıçları seçmektir. Atış poligonunda, atıcıların bazıları gerçekten nişan almak konusunda iyiydi. Aynı şey bütün profesyonel yargı görevleri için geçerlidir: En yüksek beceriye sahip olanlar hem daha az gürültülü hem de daha az ön yargılı olurlar. En iyi yargıçların nereden bulunacağı sorusunun cevabı bazen çok açıktır: Bir satranç sorununu çözmek istiyorsanız, bu kitabın yazarlarına değil, bir satranç ustasına danışmalısınız. Ancak çoğu sorunda, daha üstün yargıçların özelliklerini ayırt etmek daha zordur. Bu özellikler Bölüm 18'in konusu olacak.

Ardından yargı hatalarının azaltılmasına yaklaşımları ele alacağız. Psikolojik ön yargılar hem istatistiki ön yargı hem de gürültüyle ilişkilendirilirler. Bölüm 19'da göreceğimiz gibi, psikolojik ön yargılara, kimi açık başarısızlıklar, kimi de açık başarılarla sonuçlanan birçok karşı koyma girişimi oldu. Ön yargıyı giderme stratejilerini kısaca gözden geçirecek ve bildiğimiz kadarıyla sistematik olarak keşfedilmemiş, başarı vadeden bir yaklaşım önereceğiz: atanmış bir *karar gözlemcisinden* gerçek zamanda bir grubun çalışmasının bir ya da birkaç tanıdık ön yargıdan etkilendiğine işaret edebilecek tanısal işaretler aramasını istemek. Ek B bir karar gözlemcisinin kullanabileceği ön yargı kontrol listesine bir örnek sunuyor.

Daha sonra kitabın bu kısmının ana odağına ilerleyeceğiz: gürültüyle mücadele. İnsan yargılarındaki gürültüyü azaltmak için önerdiğimiz yaklaşıma, k*arar hijyeni* konusuna gireceğiz. Beş farklı alandan vaka incelemeleri sunacağız. Her alanda gürültünün sıklığını ve ürettiği dehşet hikâyelerinin bazılarını inceleyeceğiz. Ayrıca gürültüyü azaltma çabalarının başarısını –ve başarısızlı-

ğını– gözden geçireceğiz. Her alanda, elbette birden fazla yaklaşım kullanıldı ama sunumu kolaylaştırmak için, her bölüm tek bir karar hijyeni stratejisini vurguluyor.

Bölüm 20'de *bilgiyi sıralamanın* önemini gösteren adli tıp konusuna gireceğiz. Tutarlılık arayışı insanları eldeki kısıtlı kanıtlara dayanarak erken izlenimler oluşturmaya ve ortaya çıkan peşin hükümlerini teyit etmeye zorlar. Bu yargı sürecinin erken döneminde alakasız bilgilere maruz kalmamayı önemli kılar.

Bölüm 21'de en önemli gürültüyü azaltma stratejilerinden birinin (*çok sayıda bağımsız yargıyı bir araya toplamak*) değerini öne çıkaran bir tahmin örneğine döneceğiz. "Kalabalıkların bilgeliği" prensibi gürültüyü azaltma garantili "çoklu bağımsız yargının ortalamasını alma" işlemine dayanır. Düz ortalama almanın ötesinde, yargıları birleştirmenin yine tahmin örneğiyle resmedilen başka yöntemleri de vardır.

Bölüm 22, tıpta gürültüyü ve azaltılması yönündeki çabaları ele alıyor. Daha önce cezai hüküm örneğinde değindiğimiz bir gürültüyü azaltma stratejisinin (*yargı kılavuzu*) önemine ve genel uygulanabilirliğine dikkat çekiyor. Kılavuzlar güçlü gürültü azaltma mekanizmaları olabilirler çünkü nihai yargılarda karar vericiler arasındaki değişkenliği doğrudan azaltırlar.

Bölüm 23'te iş hayatındaki tanıdık bir zorluğa bakacağız: performans değerlendirmeleri. Bu alanda gürültüyü azaltma çabaları *dış bakış temelli ortak bir ölçek* kullanmanın kritik önemini gösteriyor. Bu, çok basit bir nedenden önemli bir karar hijyeni stratejisidir: Yargı bir izlenimin bir ölçeğe aktarılmasını gerektirir ve farklı yargıçlar farklı ölçekler kullanırsa gürültü olur.

Bölüm 24, son yüzyıldır kapsamlı olarak araştırılan alakalı ama ayrı bir konuyu, personel seçimini inceliyor. Temel bir karar hijyeni stratejisinin, *karmaşık yargıları yapılandırmanın* değerine dikkat çekiyor. *Yapılandırma* sözcüğüyle bir yargıyı ögelerine parçalamayı, veri toplama sürecini girdilerin birbirinden bağımsız

olacağı şekilde yönetmeyi, bütüncül tartışmayı ve nihai kararı bütün bu girdiler toplanana kadar geciktirmeyi kastediyoruz.

Personel seçimi alanından aldığımız derslerden yola çıkarak Bölüm 25'te seçenek değerlendirmeye *aracı değerlendirmeler protokolü (kısaca ADP)* adını verdiğimiz genel bir yaklaşım sunuyoruz. ADP "seçeneklerin adaylar gibi olduğu" önermesinden başlıyor ve yapılandırılmış karar almanın, yukarıda bahsedilen diğer karar hijyeni stratejileriyle birlikte hem tekrarlı hem tekil kararlar için tipik bir karar sürecine nasıl dâhil edilebileceğini şematik olarak gösteriyor.

Yola çıkmadan önce genel bir noktaya dikkat çekelim: Her karar hijyeni stratejisinin olası getirilerini çeşitli bağlamlarda belirlemek, hatta sayısallaştırmak değerli olurdu. Ayrıca bu stratejilerden hangilerinin en faydalı olduğunu ve nasıl karşılaştırılabileceklerini bilmek de değerli olurdu. Bilgi akışı kontrol edildiğinde, gürültü ne dereceye kadar azaltılır? Amaç gürültüyü azaltmaksa, pratikte kaç yargı bir araya toplanmalıdır. Yargıları yapılandırmak değerli olabilir ama farklı bağlamlarda tam olarak ne kadar değerlidir?

Gürültü konusuna çok az dikkat edildiği için, bu soruların ucu şimdilik açık ve araştırmalar onları zaman içinde ele alabilir. Pratik sebeplerden ötürü, stratejilerin her birinin getirileri kullanıldıkları ortama bağlı olacaktır. Kılavuzların benimsenmesini düşünün: Bazen çok büyük getirileri olacaktır (birtakım tıbbi teşhislerde göreceğimiz gibi). Ancak başka ortamlarda, kılavuz benimsemenin getirileri belki zaten çok fazla gürültü olmadığı ya da en iyi kılavuzlar bile hatayı çok fazla azaltmadığı için çok mütevazı olabilir. Herhangi bir bağlamda, yargıçlar Kısım 6'da ele alacağımız, karar hijyeni stratejilerinden sağlanacak olası faydalar ve karşılık gelen maliyetler konusunda daha net bir anlayışa ulaşmayı amaç edinmelidir.

BÖLÜM 18

Daha İyi Yargılar için Daha İyi Yargıçlar

Buraya kadar, aralarında ayrım yapmadan çoğunlukla insan yargıçlardan bahsettik. Ancak yargı gerektiren her görevde, bazı insanların diğerlerinden daha iyi performans sergileyeceği çok açıktır. Kalabalıkların bilgeliğine dayanan bir yargı kümesi bile, söz konusu kalabalığın daha becerikli insanlardan oluşması hâlinde daha iyi olacaktır. Bu durumda, bu daha iyi yargıçların nasıl saptanacağı, önemli bir sorudur.

Üç şey önemlidir. Kararlar, onları alanlar iyi eğitimli, daha zeki ve doğru bilişsel tarza sahip olduğunda hem daha az gürültülü hem de daha az ön yargılı olur. Başka bir ifadeyle, iyi yargılar, bildiklerinize, ne kadar iyi düşündüğünüze ve *nasıl* düşündüğünüze bağlıdır. İyi yargıçlar tecrübeli ve akıllı olma eğilimindedirler ama aynı zamanda, aktif olarak açık görüşlü ve yeni bilgilerden ders almaya isteklidirler.

Uzmanlar ve Saygı Uyandıran Uzmanlar

Yargıçların becerilerinin yargılarının kalitesini etkilediğini söylemek neredeyse gereksiz. Örneğin, becerikli radyologların zatürreyi doğru olarak teşhis etmeleri daha olasıdır ve dünya olaylarının tahmin edilmesinde, süper olmayan meslektaşlarını güvenilir bir şekilde geride bırakan "süper tahminciler" vardır. Hukukun belli bir alanında gerçek uzmanlar olan bir grup hukukçuyu bir araya toplarsanız mahkemedeki yaygın hukuki anlaşmazlıkların sonucu hakkında benzer ve iyi öngörülerde bulunacaklardır. Beceri seviyesi yüksek insanlar daha az gürültülüdür ve daha az ön yargı sergilerler.

Bu insanlar söz konusu görevlerde gerçek uzmanlardır. Başkalarına üstünlükleri sonuç verilerinin ulaşılabilirliği sayesinde doğrulanabilir. En azından prensipte bir doktoru, tahminciyi ya da hukukçuyu geçmişte hangi sıklıkla haklı çıktığına göre seçebiliriz. (Bilinen nedenlerden ötürü, bu yaklaşım pratikte zor olabilir, aile hekiminizi bir uzmanlık sınavına tabii tutmayı denemenizi önermeyiz.)

Ayrıca belirttiğimiz gibi, pek çok yargı doğrulanamaz. Belirli sınırlar içinde, yargıların hedef aldığı gerçek değeri kolayca bilemez ya da tartışmaya yer bırakmayacak şekilde tanımlayamayız. Sigortalama ve cezai hükümler bu kategoriye girer, tıpkı şarap tadımı, kompozisyon notlama, kitap ve film puanlama ve sayısız başka yargılar gibi. Bununla birlikte, bu alanlardaki bazı profesyoneller uzman olarak adlandırılabilir. Bu uzmanların yargılarına duyduğumuz güven tamamen meslektaşlarından gördükleri saygıya dayalıdır. Onlara *saygı uyandıran uzmanlar* deriz.

Saygı uyandıran uzman terimi saygısızlık niyetini taşımıyor. Bazı uzmanların yargılarının isabetliliğinin değerlendirmeye tabi tutulmaması bir eleştiri değil; pek çok alanda hayatın gerçeği. Pek çok üniversite hocası, bilim insanı ve yönetim danışmanı saygı uyandıran uzmandır. İnandırıcılıkları öğrencilerinin, mes-

lektaşlarının ya da müşterilerinin saygısına dayanır. Bütün bu alanlarda ve daha pek çoğunda, bir profesyonelin yargıları sadece emsallerininkilerle karşılaştırılabilir.

Kimin haklı, kimin haksız olduğunu belirleyecek gerçek değerlerin yokluğunda, saygı uyandıran uzmanların görüşüne birbirlerine ters düştüklerinde bile değer veririz. Örneğin, birden fazla siyaset analizcisinin bir diplomatik krizin nedeni ve sonrasında olacaklar konusunda keskin bir şekilde farklı bakış açıları ifade ettikleri bir panel düşünün. (Bu anlaşmazlık, sıra dışı değildir, hepsi hemfikir olsalardı çok ilginç bir panel olmazdı.) Analizcilerin hepsi bir doğru bakış açısı olduğuna ve kendilerininkinin ona en yakın bakış açısı olduğuna inanır. Dinledikçe, analizcilerden birkaçının eşit derecede etkileyici ve argümanlarının eşit derecede ikna edici olduğunu görebilirsiniz. O zaman içlerinden hangisinin haklı olduğunu bilemezsiniz (ve hatta analizleri açıkça doğrulanabilir tahminler gibi formüle edilmediğinde, daha sonra bile bilemeyebilirsiniz). Analizcilerin en azından birkaçının yanıldığını bilirsiniz, çünkü hemfikir değildirler. Yine de uzmanlıklarına saygı duyarsınız.

Ya da tahminde bulunmayan başka bir uzman grubu düşünün. Her biri iyi eğitimli üç ahlak felsefecisi bir odada toplanmış olsun. İçlerinden bir Immanuel Kant'ı, diğeri Jeremy Bentham'ı, üçüncüsü ise Aristo'yu takip ediyor. Ahlaklılığın neyi gerektirdiği konusunda, apayrı görüşlere sahipler. Mesele, yalan söylemenin meşru olup olmadığı ve ne zaman meşru olduğu ya da hayvan hakları veya suçluları cezalandırmanın amacı olsun. Dikkatle dinlersiniz. Düşüncelerinin berraklığına ve netliğine hayran kalabilirsiniz. Filozoflardan biriyle hemfikir olma eğilimindesinizdir ama hepsine saygı duyarsınız.

Bunu neden yaparsınız? Daha genel anlamda, kendileri yargılarının kalitesi için saygı gören insanlar ortada uzmanlığı objektif olarak doğrulayacak bir veri olmadığında, bir kişinin uzmanlığına

güvenmeye nasıl karar verirler? Bir kişiyi saygı uyandıran uzman yapan nedir?

Cevabın bir kısmını paylaşılan normlar ya da profesyonel doktrin oluşturuyor. Uzmanlar genellikle profesyonel topluluklar tarafından onaylanır, örgütlerinden eğitim ve denetim alır. Uzmanlıklarını tamamlayan doktorlar ve kıdemli bir ortak tarafından eğitilen genç avukatlar, sadece mesleklerinin teknik araçlarını öğrenmezler: Belli yöntemleri kullanmak ve belli normlara uymak için eğitilirler.

Ortak normlar profesyonellere hangi girdilerin hesaba katılması gerektiği, nihai kararlarını nasıl verecekleri ve nasıl gerekçelendirecekleri konusunda bir his sağlar. Örneğin sigorta şirketinde, sigorta eksperleri bir talebi değerlendirmek için bir kontrol listesine dâhil edilmesi gereken ilgili hususlar konusunda anlaşmak ve onları tanımlamak konusunda hiç sıkıntı yaşamadılar.

Bu görüş birliği elbette sigorta eksperlerini, hasar talep değerlendirmelerinde büyük farklılıklar sergilemekten alıkoymadı çünkü doktrin nasıl ilerleneceğini tam olarak detaylandırmaz. Mekanik bir şekilde uyulacak bir tarif değildir. Doktrin yoruma yer bırakır. Uzmanlar hâlâ hesaplama değil, yargı üretirler. Bu yüzden gürültünün olması kaçınılmazdır. Aynı eğitimi almış ve uyguladıkları doktrin konusunda hemfikir olan profesyoneller bile, uygulama sırasında birbirlerinden ayrı düşerler.

Ortak norm bilgisinin ötesinde, tecrübe de gereklidir. Uzmanlık alanınız satranç, konser piyanistliği ya da cirit atmaysa genç bir dâhi olabilirsiniz çünkü sonuçlar performans seviyenizi teyit eder. Ancak risk değerlendirme uzmanları, parmak izi inceleme uzmanları ya da yargıçlar, inanılırlık için genellikle birkaç yıllık tecrübeye ihtiyaç duyarlar. Sigortalamada genç dâhiler yoktur.

Saygı uyandıran uzmanların bir diğer özelliği de yargılarını güvenle verme ve açıklama becerileridir. Kendine güvenen insanlara, şüphelerini belli edenlerden daha çok güveniriz. Güven

kısayolu bir grupta öz güvenli insanların, öz güvenli olmak için bir nedenleri olmasa bile, diğerlerinden daha fazla ağırlık sahibi olduğuna işaret eder. Saygı uyandıran uzmanlar tutarlı hikâyeler inşa etmekte ustalaşırlar. Tecrübeleri örüntüleri tanımalarını ve önceki vakalarla benzerliğe göre akıl yürütmelerini ve hızlı bir şekilde hipotez oluşturup onaylamalarını sağlar. Gördükleri olguları öz güven uyandıran tutarlı bir hikâyeye kolayca sığdırırlar.

Zekâ

Eğitim, tecrübe ve öz güven saygı uyandıran uzmanlara güven uyandırma imkânı verir. Ancak bu özellikler yargılarının kalitesini garantilemez. Hangi uzmanların daha iyi karar vermeye yatkın olduğunu nasıl bilebiliriz?

Genel zekânın daha iyi yargıyla bağdaştırılabileceğine inanmak için çok sebep var. Zekâ neredeyse her alanda iyi performansla bağıntılıdır. Diğer her şey eşit olduğunda, sadece daha yüksek akademik başarıyla değil, aynı zamanda daha yüksek iş performansıyla da ilişkilendirilir.

Zekâ veya genel zihinsel yetenek ölçümleriyle (artık zekâ katsayısı yani IQ yerine kullanılması tercih edilen terim, kısaca GZY) ilgili tartışmalarda pek çok karşı görüş ve yanlış anlaşılmalar doğar. Zekânın doğasıyla ilgili birtakım yanlış yorumlamalar hâlâ sürmektedir; aslında testler kısmen kalıtım özelliklerinin bir işlevi olan ve kısmen eğitim fırsatlarının dâhil olduğu çevre şartlarından etkilenen geliştirilmiş becerileri ölçer. Pek çok insan ayrıca tanımlanabilir sosyal gruplarda GZY'ye dayalı seçmenin olumsuz etkisi ve GZY testlerinin seçme amacıyla kullanımının meşruluğu konusunda kaygılar taşımaktadır.

Testlerin kullanımıyla ilgili bu kaygıları, kestirimci değerlerinin gerçekliğinden ayırmalıyız. ABD ordusu bir asır önce zihinsel beceri testlerini kullanmaya başladığından beri, bilişsel test puanları ve

sonraki performans arasındaki bağlantıyı ölçen binlerce çalışma yapıldı. Bu çalışma yığınından çıkan mesaj ise muğlak. Bir incelemenin ifade ettiği gibi, "GZY hem ulaşılan mesleki seviyeyi hem de kişinin seçtiği meslekteki performansını öngörüyor ve bunu beceri, özellik, yaradılış ve iş tecrübesi gibi diğer kriterlerden daha iyi yapıyor." Elbette başka bilişsel beceriler de önemli (bu konuya daha sonra döneceğiz). Ve sorumluluk ve "dayanıklılık ve uzun vadeli hedeflerin peşinden gitme tutkusu" olarak tanımlanan azim gibi pek çok kişilik özelliği. Ve evet, pratik zekâ ve yaratıcılık gibi, GZY testlerinin ölçmediği birçok zekâ türü var. Psikologlar ve nörobilimciler kristalize zekâyla (aritmetik işlemler dâhil, dünya hakkında depolanmış bilgilerden faydalanarak sorun çözme becerisi) akıcı zekâyı (yeni sorunları çözme becerisi) birbirinden ayrı tutuyorlar.

Bununla birlikte sözel, sayısal ve uzaysal problemlerle ilgili sorular içeren standardize testlerle ölçülen GZY, hamlığına ve sınırlarına rağmen, önemli sonuçların en iyi kestiricisi olma özelliğini açık ara koruyor. Az önce bahsettiğimiz incelemede eklendiği gibi, GZY'nin kestirimci gücü "psikolojik araştırmalarda bulunanların çoğundan daha büyük." Genel zihinsel beceri ve iş başarısı arasındaki ilişkinin gücü, söz konusu işin karmaşıklığının artmasıyla paralel ve mantıklı bir şekilde artıyor. Roket bilimciler için zekâ, basit görevleri icra edenler için olduğundan çok daha önemli. Karmaşıklık düzeyi daha yüksek olan işlerde, standardize test puanları ve iş performansları arasında gözlemlenebilen korelasyonlar 0,50 (YU = %67) diliminde. Daha önce belirttiğimiz gibi, 0,50'lik bir korelasyon sosyal bilim standartlarında çok güçlü bir kestirimci değeri işaret eder.

Özellikle beceri gerektiren profesyonel yargılarla ilgili tartışmalarda, zekâ ölçümlerinin alakasıyla ilgili önemli ve sık rastlanan bir itiraz, bu tür yargıları üreten kişilerin hepsinin muhtemelen yüksek GZY sahibi bireyler olduğu. Doktorlar, yargıçlar ya da

kıdemli risk değerlendirme uzmanları nüfusun genelinden çok daha iyi eğitimlidirler ve her tür bilişsel beceri ölçümünde daha yüksek puanlara ulaşmaları kuvvetle muhtemeldir. Akla yatkın bir şekilde, yüksek GZY'nin aralarında çok az fark yarattığına, GZY'nin o havuz içindeki başarı farklarının kaynağı değil, sadece yüksek başarı havuzuna giriş bileti olduğuna inanabilirsiniz.

Ancak bu yaygın inanç doğru değildir. Hiç şüphesiz, belirli bir meslekte bulunan GZY'lerin aralığı meslekler sıralamasının en altında, en üstünde olduğundan daha geniştir. Daha düşük seviye mesleklerde yüksek GZY'li bireyler vardır ama avukatların, kimyagerlerin ya da mühendislerin arasında ortalama GZY'nin altında kalan neredeyse kimseyi bulamazsınız. Bu bakış açısıyla, yüksek zihinsel yeteneğin yüksek statülü mesleklere girebilmek için gerek şart olduğu açıktır.

Ancak bu ölçüm bu grupların *içindeki* başarı farklarını yakalamakta yetersiz kalmaktadır. Bilişsel beceri ölçümüne göre (13 yaşında değerlendirilmiş) en üst %1'lik dilimdeki insanların arasında bile, olağanüstü sonuçlar GZY ile yakından ilişkilidir. Bu en üst %1'lik dilimin alt çeyreğinde yer alanlarla karşılaştırıldıklarında, üst çeyrektekilerin doktora seviyesine ulaşma, kitap yayımlama ya da patent alma ihtimali iki ya da üç kat yüksektir. Başka bir deyişle GZY'deki fark sadece %1'lik dilimle %20'lik ve %50'lik dilimler arasında fark yaratmaz, %1'lik dilimin ilk %12'siyle, son %13'ü arasında da fark yaratır, hem de çok!

Yetenek ve sonuçları arasındaki bağlantının bir başka çarpıcı sunumunda, 2013 yılında yapılan bir çalışma Fortune 500 şirketlerinin CEO'larına ve 424 Amerikalı milyardere odaklandı (Amerikalıların servet açısından ilk % 0,00001'lik dilimi). Tahmin edileceği gibi, bu hiper elit grupların entelektüel açıdan en becerikli insanlar arasından çekilmiş kişilerden oluştuğu ortaya kondu. Ancak çalışma aynı zamanda, bu grupların *içinde* yüksek eğitim ve beceri seviyelerinin daha yüksek telafi (CEOlar için)

ve net değerle (milyarderler için) alakalı olduğunu da saptadı. Yeri gelmişken, Steve Jobs, Bill Gates ve Mark Zuckerberg gibi üniversiteyi yarım bırakmış ünlü milyarderler, ormanı gizleyen ağaçlardır. Amerikalı yetişkinlerin yaklaşık üçte biri üniversite diplomasına sahipken, bu oran milyarderlerde %88'e çıkıyor.

Sonuç çok net. GZY yargı gerektiren mesleklerde, yüksek beceri sahibi adaylardan oluşan bir havuz içinde bile, performans kalitesine önemli ölçüde katkı sağlıyor. GZY'nin fark yaratmaz düzeye geldiği bir eşik olduğu kavramı kanıtlarla desteklenmiyor. Bu çıkarım ayrıca profesyonel yargılar doğrulanamaz olduğunda ama görünmez boğa gözüne eriştikleri varsayıldığında, yüksek beceri sahibi insanların hedefe yaklaşma olasılığının daha yüksek olduğunu ortaya koyuyor. Karar alacak insanlar seçmeniz gerektiğinde en yüksek zihinsel beceriye sahip olanları seçmek çok mantıklı görünüyor.

Ancak bu akıl yürütme çizgisinin önemli bir sınırı var. Herkesi standardize testlerden geçiremeyeceğinize göre, en yüksek GZY'li insanların kimler olduğunu tahmin etmeniz gerekecek. Ve yüksek GZY haklı olduğunuza başkalarını ikna etmenin de aralarında olduğu pek çok cephede performansı artırır. Yüksek zihinsel beceri sahibi insanların daha iyi kararlar verme ve gerçek uzmanlar olma olasılığı diğerlerinden yüksektir ama aynı zamanda emsallerini etkileme, başkalarının güvenini kazanma ve gerçeklik geri beslemesinin olmadığı durumlarda saygı uyandıran uzmanlar olmaları da daha olasıdır. Orta Çağ astrologları kendi çağlarının en yüksek GZY'li insanları arasında olmalıydılar.

Zeki görünen ve zekice konuşan, yargıları için ikna edici bir mantığı dile getirebilen insanlara güvenmek akla yatkındır ancak bu strateji yetersizdir ve geri bile tepebilir. Öyleyse, gerçek uzmanları ayırt etmenin başka yolları var mıdır? En iyi yargıları üretebilen insanların ayırt edilebilir başka özellikleri de var mı?

Bilişsel Tarz

Zihinsel beceriden bağımsız olarak, insanlar *bilişsel tarzlarında* ya da yargı görevlerine yaklaşımlarında farklılık gösterirler. Bilişsel tarzları kavramak için pek çok araç geliştirilmiştir. Bu ölçümlerin çoğu GZY ile (ve birbirleriyle) bağlantılıdır ama farklı şeyleri ölçerler.

Bu tür ölçümlerden biri, artık her yerde karşımıza çıkan şu top ve sopayla ilgili soruyla ünlenen *bilişsel yansıma* testidir (BYT). "Bir sopa ve top toplamda 1,10 dolar tutuyor. Sopanın fiyatı toptan 1 dolar yüksek. Topun fiyatı nedir?" Bilişsel yansımayı ölçmek için sorulan diğer sorulardan biri de şudur: "Bir yarışta koşuyorsanız ve ikinci sıradaki kişiyi geçerseniz, kaçıncı sıradasınızdır?" BYT soruları insanların ilk akla gelen (yanlış) cevabı (sopa ve top sorusu için "10 sent" ve yarış örneği için "birinci") aşma olasılığını ölçmeye çalışır. Daha düşük BYT puanları aralarında hayaletlere, astrolojiye ve duyu ötesi algıya inanmanın da olduğu birçok gerçek dünya yargısı ve inancıyla bağlantılıdır. Puanlar insanların açıkça yanlış "sahte haberlere" kanıp kanmayacağını da öngörür. Hatta insanların akıllı telefonlarını ne kadar çok kullandıklarıyla da ilişkilidirler.

BYT pek çok kişi tarafından daha geniş bir kavramı, dürtüsel düşünce sürecine karşı yansıtıcı düşünce süreci kullanma eğilimini ölçen bir alet olarak görülür. Basit bir ifadeyle, bazı insanlar dikkatli bir düşünme şeklini benimserken, diğerleri aynı sorun karşısında ilk dürtülerine güvenme eğilimdedir. Bizim terminolojimizle BYT insanların hızlı Sistem 1 düşünme yerine, yavaş Sistem 2 düşünmeye yatkınlığının ölçümü olarak görülebilir.

Bu eğilimi ölçmek için başka öz değerlendirme yöntemleri de geliştirildi (ve elbette bütün bu testler kendi aralarında bağlantılıdır). Örneğin biliş ihtiyacı ölçeği insanlara sorunlar hakkında ne kadar çok düşünmekten hoşlandıklarını sorar. Ölçekte yüksek bir puan almak için "Sadece hatırı sayılır miktarda zihinsel çaba harcayarak ulaşılabilen hedefler belirleme eğilimindeyim"

görüşüne katılmalı ve "Düşünmek eğlence anlayışıma uymuyor" görüşüne karşı olmalısınız. Yüksek biliş ihtiyacı olan insanlar bilinen bilişsel ön yargılara daha az açık olma eğilimindedirler. Bildirilen daha tuhaf ilişkilendirmeler de var: Spoiler uyarısı içeren film eleştirilerinden kaçınıyorsanız, muhtemelen daha yüksek biliş ihtiyacınız vardır, biliş ihtiyacı ölçeğinde daha aşağıda olan insanlar spoiler'la "bozulmuş" hikâyeleri tercih ederler.

Bu öz değerlendirmeye dayanan bir ölçek olduğu ve sosyal açıdan arzulanan cevap hayli bariz olduğu için, ölçek akla birtakım soruları getiriyor. İnsanları etkilemeye çalışan bir kişinin, "Düşünmek eğlence anlayışıma uymuyor" beyanını desteklemesi pek olası değil. Bu nedenle başka testler öz tanımlamaları kullanmak yerine becerileri ölçmeye çalışırlar.

Bu testlerin bir örneği insanların karar alırken aşırı öz güven ya da risk algılarında tutarsızlık gibi tipik hatalar yapmaya ne kadar yatkın olduğunu ölçen Yetişkin Karar Alma Kabiliyeti ölçeğidir. Bir diğeri hem akılcı düşünmeye yatkınlığı hem de birtakım öğrenilebilir becerileri içeren, eleştirel düşünme becerilerine odaklanan Halpern Eleştirel Düşünme Değerlendirmesi'dir. Bu değerlendirme sırasında size şu tür sorular sorulur: "Bir arkadaşınızın iki kilo verme programından hangisini seçeceği konusunda sizden tavsiye istediğini düşünün. Programlardan biri müşterilerin ortalama on bir kilo verdiğini, diğeri ise ortalama on üç buçuk kilo verdiklerini bildiriyor. Bu programlardan birini seçmek için hangi soruların cevaplanmasını isterdiniz?" Örneğin cevap olarak kaç kişinin bu kadar kilo verdiğini ve bu kilo kaybını bir yıl ya da daha uzun süre koruyup korumadıklarını bilmek istediğinizi söylerseniz, eleştirel düşünme uyguladığınız için puan alırsınız. Yetişkin Karar Alma Kabiliyeti ölçeğinde ya da Halpern değerlendirmesinde daha yüksek puanlar alan insanlar hayatta daha iyi kararlar verir gibi görünüyorlar: Kiralık bir filmi geç iade ettikleri için ceza ödemek ya

da istenmeyen bir hamilelik gibi, kötü tercihlerin neden olduğu tatsız hayat tecrübelerini daha az yaşıyorlar.

Bütün bu bilişsel tarz ve beceri ölçümlerinin –ve daha pek çoğunun– genel olarak yargıyı öngördüğünü varsaymak mantıklı görünüyor. Öte yandan alakaları göreve göre değişiyor gibi görünüyor. Uriel Haran, Ilana Ritov ve Barbara Mellers tahmin becerisini öngörebilecek bilişsel tarzları araştırırken, biliş ihtiyacının ek bilgi için kimin daha fazla çaba harcayacağını öngörmediğini saptadılar. Ayrıca, biliş ihtiyacının daha yüksek performansla güvenilir bir şekilde bağlantılı olmadığı sonucuna da vardılar.

Buldukları, tahmin performansını öngören tek bilişsel tarz ya da kişilik özelliği, psikoloji profesörü Jonathan Baron'ın "aktif olarak açık görüşlü düşünme"yi ölçmek için geliştirdiği bir başka ölçek oldu. Aktif olarak açık görüşlü olmak önceden var olan hipotezlerinizle çelişen bilgiyi aktif olarak aramak anlamına gelir. Bu tür bilgiler başka insanların karşıt görüşlerini ve yeni kanıtların eski inançlara karşı dikkatli bir şekilde tartılarak değerlendirilmesini içerir. Aktif olarak açık görüşlü insanlar: "Kişinin karşı bir görüş tarafından ikna edilmesine izin vermesi iyi karakter işaretidir" gibi beyanlara katılırlar. "Fikrinizi değiştirmenin bir zayıflık işareti olduğu" ya da "sezgi karar almakta en iyi rehberdir" gibi önermelere itiraz ederler.

Başka bir deyişle, bilişsel yansıtma ve biliş ihtiyacı puanları yavaş ve dikkatli düşünmeye yatkınlığı ölçerken, aktif olarak açık görüşlü düşünme bunun ötesine geçer. Bu, yargılarının tamamlanmamış bir çalışma olduğunun sürekli olarak farkında olan ve düzeltilmek için yanıp tutuşanların tevazusudur. Bölüm 21'de bu düşünme tarzının sürekli olarak fikir değiştiren ve yeni bilgiye cevap olarak inançlarını değiştiren en iyi tahmincilerin baş özelliği olduğunu göreceğiz. İlginç bir şekilde, fiilen açık görüşlü olmanın öğretilebilir bir beceri olduğu yönünde kanıtlar mevcut.

Burada amacımız belli bir alanda iyi kararlar alacak bireylerin nasıl seçileceği konusunda sert ve hızlı sonuçlara varmak değil. Ama bu kısa incelemeden iki genel prensip doğuyor: Öncelikle, uzmanlığın doğru değerlerle karşılaştırmayla teyit edilebileceği alanlarla (hava durumu tahmini gibi) saygı uyandıran uzmanların yetki bölgesi olan alanlar arasındaki farkları ayırt etmek akılcı olacaktır. Bir politika analizcisi güzel ve ikna edici konuşuyor olabilir ve bir satranç ustası utangaç görünebilir ve hamlelerinin bazılarının arkasındaki mantığı açıklamayı beceremeyebilir. Ancak, politika analizcisinin profesyonel yargısını satranç ustasınınkinden daha büyük bir şüphecilikle ele almalıyız.

İkinci olarak bazı yargıçlar eşit derecede kalifiye ve tecrübeli emsallerinden daha iyi olacaklardır. Daha iyilerse, ön yargılı ya da gürültülü olma olasılıkları daha düşüktür. Bu farklılıkları açıklayan pek çok şeyin arasında, zekâ ve bilişsel tarz önemlidir. Her ne kadar hiçbir ölçü ya da ölçek yargı kalitesini tek başına şüpheye yer bırakmadan öngöremese de aktif olarak daha önceki inançlarıyla çelişebilecek yeni bilgilerin arayışında olan, bu bilgiyi mevcut bakış açılarına entegre etmekte sistemli hareket eden ve sonucunda fikirlerini değiştirmeye istekli hatta hevesli olan insanları yeğleyebilirsiniz.

Kusursuz yargı becerisi olan insanların kişiliği genel olarak kabul gören kararlı lider stereotipine uymayabilir. İnsanlar genellikle kararlı, açık ve doğru olanı hızlı bir şekilde ve iliklerinde duyar gibi görünen liderlere güvenir ve onları sever. Bu tür liderler güven uyandırır. Ancak kanıtlar gösteriyor ki amaç hatayı azaltmaksa, liderlerin (ve başkalarının) karşı argümanlara açık olması ve yanılıyor olabileceklerini bilmesi daha iyidir. Kararlı olacaklarsa bu sürecin başında değil, en sonunda olmalıdır.

Daha İyi Yargıçlardan Bahsederken

"Uzmansınız. Ama yargılarınız doğrulanabilir mi, yoksa bir saygı uyandıran uzman mısınız?"

"İki görüş arasında seçim yapmamız gerek ve bu bireylerin uzmanlığı ve geçmiş performansları konusunda bir şey bilmiyoruz. Daha zeki olanın tavsiyesine uyalım."

"Zekâ hikâyenin sadece bir kısmıdır. İnsanların *nasıl* düşündüğü de önemlidir. Belki de en akıllı yerine, en düşünceli ve en açık görüşlü kişiyi seçmeliyiz."

BÖLÜM 19

Ön Yargıları Gidermek ve Karar Hijyeni

Pek çok araştırmacı ve örgüt yargıları ön yargılardan arındırmayı amaç edinmiştir. Bu bölümde onların merkez bulgularını ele alacağız. Farklı ön yargı giderme girişimlerini karşılaştıracak ve daha fazla incelenmeyi hak eden bir müdahalenin üstünde duracağız. Daha sonra gürültünün azaltılması konusuna dönecek ve karar hijyeni düşüncesini sizlerle paylaşacağız.

Sonrasında ve Öncesinde Ön Yargıyı Gidermek

Ön yargıyı gidermeye başlıca iki yaklaşımı tanımlamanın iyi bir yolu, ölçüm benzetmesine dönmek olacaktır. Banyo tartınız kilonuza ortalama bir kilo ekliyor olsun. Tartınız ön yargılıdır. Ama bu onu işe yaramaz kılmaz. Ön yargısını iki olası yoldan biriyle ele alabilirsiniz: Kötü kalpli tartınızın her ölçümünü bir kilo eksilterek düzeltebilirsiniz. Bu biraz yorucu olabilir (ve yapmayı unutabilirsiniz). Kadranı ayarlamak ve aletin doğruluğunu temelli olarak iyileştirmek bir alternatif olabilir.

Ölçümlerde ön yargıyı gidermeye bu iki yaklaşım yargılarda ön yargıyı giderme müdahaleleriyle direkt benzerlik taşır. Bu müdahaleler, kararları verildikten sonra düzelterek "sonrasında" ya da bir yargı veya karar öncesinde müdahale ederek "öncesinde" devreye girerler.

"Sonrasında" ya da düzeltici ön yargı giderme genellikle sezgisel olarak yapılır. Bir projeden sorumlu bir ekibe gözetmenlik yaptığınızı ve takımın projeyi üç ayda tamamlayabileceğini tahmin ettiğini varsayalım. Takım üyelerinin yargısına bir tampon ekleyip dört ay ya da daha uzun sürelik bir planlama yapmak ve dolayısıyla var olduğunu varsaydığınız bir ön yargıyı düzeltmek (planlama yanılgısı) isteyebilirsiniz.

Ön yargı düzeltmenin bu çeşidi bazen daha sistematik şekilde ele alınır. İngiltere'de Majestelerinin Hazinesi, program ve projelerin nasıl değerlendirilmesi gerektiği konusunda rehber olarak *The Green Book*'u yayımladı. Rehber, iyimser ön yargıları gidermek için planlayıcıları projelerin maliyet ve süre tahminlerinde yüzde ayarlamaları yapmaya teşvik ediyor. Bu ayarlamalar ideal olarak söz konusu örgütün tarihteki iyimserlik ön yargısı seviyelerine dayandırılmalı. Böyle bir tarihsel veriye erişilemeyen durumlarda, *The Green Book* her proje türüne özel ayar yüzdeleri uygulamayı tavsiye ediyor.

"Öncesinde" yapılan ya da önleyici ön yargı giderme müdahaleleri ise iki geniş kategoriye ayrılıyor. En umut vadedenler yargıya varılan ya da kararın alındığı ortamı değiştirecek şekilde tasarlanıyor. Bu tür modifikasyonlar ya da *dürtmeler* ön yargıların etkisini azaltmayı ve hatta daha iyi bir karar üretmek için ön yargılardan yardım almayı hedefler. Basit bir örneği emeklilik planlarına otomatik katılımdır. Ataleti, ertelemeyi ve iyimser ön yargıyı yenmek için tasarlanan otomatik katılım, çalışanların sistemden çıkmayı kasten seçmemeleri durumunda, emeklilik için tasarruf yapmalarını sağlar. Otomatik katılım iştirak oranlarını

artırmada son derece etkili olmuştur. Programa bazen çalışanların gelecekteki maaş artışlarının belli bir yüzdesini tasarrufa tahsis etmeye onay verebildiği Save More Tomorrow (Yarın Daha Fazla Biriktir) planları da eşlik etmektedir. Otomatik katılım pek çok yerde, örneğin yeşil enerjiye otomatik katılımda, yoksul çocuklar için ücretsiz okul yemek planlarında ve diğer birçok yardım programında kullanılabilir.

Başka dürtmeler de seçim mimarisinin farklı yönlerinde işe yarayabilir. Örneğin, zihin sağlığı sorunlarında yardıma erişimde idari yükleri azaltarak doğru kararı kolay karar kılabilirler. Ya da bir ürünün ya da faaliyetin belli özelliklerini belirgin kılarak, mesela önceden gizli olan ücretleri aleni ve açık kılabilirler. Marketler ve web sayfaları kolayca insanları ön yargılarını yenecek bir şekilde dürtmek için tasarlanabilir. Sağlıklı yiyecekler öne çıkan yerlere konursa, insanların onları alma ihtimali artar.

Ön yargıyı öncesinde gidermenin farklı bir yolu da karar alıcıları kendi ön yargılarını fark edip yenecek şekilde eğitmektir. Bu müdahalelerin bazılarına *güçlendirici* denir; kişilerin kapasitelerini örneğin istatistik okuryazarlığı öğreterek iyileştirmeyi amaçlarlar.

İnsanları ön yargılarını yenmeye eğitmek onurlu bir girişimdir ama göründüğünden daha zorludur. Elbette eğitim faydalıdır. Örneğin, yıllarca ileri düzey istatistik dersi almış insanların istatistiksel akıl yürütmede hata yapma olasılığı daha azdır. Ancak insanlara ön yargılardan kaçınmayı öğretmek zordur. Onlarca yıllık araştırma uzmanlık alanlarında ön yargılardan kaçınmayı öğrenmiş profesyonellerin, öğrendiklerini farklı alanlara uygulamakta sıklıkla güçlük yaşadıklarını göstermektedir. Örneğin hava durumu tahmincileri, öngörülerinde aşırı öz güvenden kaçınmayı öğrenmişlerdiler. %70'lik bir yağmur şansı duyurduklarında, genellikle yüz seferden yetmişinde yağmur yağar. Bununla birlikte, genel bilgi soruları karşısında onlar da diğer insanlar kadar aşırı öz güvenli davranabilirler. Bir ön yargıyı yenmeyi öğrenmenin

zorluğu yeni bir sorunun başka bir yerde daha önce gördüğümüz bir soruna benzer olduğunu ve bir yerde gördüğünüz bir ön yargının başka yerlerde de karşımıza çıkabileceğini ayırt edebilmektir.

Araştırmacılar ve eğitimciler bu ayırdı kolaylaştırmak için geleneksel olmayan öğretme yöntemlerinde bir miktar başarı sağladılar. Bir çalışmada, Boston Üniversitesi'nden Carey Morewedge ve çalışma arkadaşları eğitici videolar ve "ciddi oyunlar"dan faydalandılar. Katılımcılar onay ön yargılarının, çapalamanın ve diğer psikolojik ön yargıların sebep olduğu hataları tanımayı öğrendiler. Her oyundan sonra, yaptıkları hatalar hakkında geri besleme aldılar ve onları tekrarlamaktan nasıl kaçınacaklarını öğrendiler. Oyunlar (ve daha düşük ölçüde videolar) katılımcıların hemen akabinde ve sekiz hafta sonra yapılan ve benzer soruların sorulduğu testlerde yaptıkları hataların sayısını azalttı. Ayrı bir çalışmada ise, Anne-Laure Sellier ve meslektaşları, eğitici bir video oyunu oynarken onay ön yargısını yenmeyi öğrenen MBA öğrencilerinin, burada öğrendiklerini bir başka derste bir iş vakasını çözerken uyguladığını gördüler. Öğrenciler bunu iki egzersiz arasında bir bağlantı olduğunun söylenmemesine rağmen yaptılar.

Ön Yargıların Giderilmesinin Sınırı

İster sonrasında ön yargıları düzeltsinler ister dürtme ve güçlendirme aracılığıyla etkilerini önlesinler, çoğu ön yargıyı giderme yaklaşımının ortak noktası belirli ve var olduğunu varsaydıkları bir ön yargıyı hedef almalarıdır. Bu genellikle akla yatkın varsayım zaman zaman yanlış çıkabilmektedir.

Bir kez daha proje planlama örneğini ele alalım. Mantıklı bir yaklaşımla, aşırı öz güvenin proje ekiplerini genel olarak etkilediğini varsayabilirsiniz ama bunun belli bir proje ekibini etkileyen tek ön yargı (hatta başlıcası) olduğundan emin olamazsınız. Belki ekip lideri benzer bir projeyle kötü bir tecrübe yaşamış ve bu nedenle

tahmin yaparken özellikle tutucu olmayı öğrenmiştir. Dolayısıyla ekip düzeltmeniz gerekeceğini düşündüğünüzün tam tersi bir hata sergileyebilir. Ya da belki de ekip tahminini benzer başka bir projeye benzerlikle geliştirmiştir ve o projenin gerçekleştirildiği süreye çapalanmıştır. Ya da belki de proje ekibi tahminlerine sizin bir tampon eklemenizi beklediği için, önerisini gerçek inancından daha iyimser kılarak ayarlamanızı etkisiz kılmıştır.

Ya da bir yatırım kararını düşünün. Yatırımın geleceği konusunda aşırı öz güven söz konusu olabilir ama bir diğer güçlü ön yargı olan kayıptan kaçınma, tam tersi etkisiyle karar vericilerin başlangıçtaki harcamalarını kaybetme riskini almaktan hoşlanmamalarına neden olabilir. Ya da birçok projede kaynak tahsis eden bir şirket düşünün. Karar vericiler yeni girişimlerin etkisi konusunda iyimser (yine aşırı öz güven) ve var olan birimlerden kaynak yönlendirmek konusunda çok çekingen olabilirler (adından da anlaşılacağı gibi, bir şeyleri olduğu gibi bırakma tercihi anlamına gelen *statüko ön yargısının* neden olduğu bir problem).

Bu örneklerin de işaret ettiği gibi, bir yargıyı tam olarak hangi psikolojik ön yargıların etkilediğini bilmek zordur. Karmaşıklık içeren her durumda, birden fazla psikolojik ön yargı devrede olup aynı yönde hata eklemek için iş birliği yapıyor ya da öngörülemez sonuçlarla birbirlerini dengeliyor olabilirler.

Kısaca, yargının sonrasında ya da öncesinde ön yargıları gidermek (belirli psikolojik ön yargıları sırasıyla düzeltmek veya önlemek) bazı durumlarda faydalıdır. Bu yaklaşımlar hatanın genel istikameti bilindiğinde ve kendini net bir istatistiksel ön yargı olarak gösterdiği durumlarda işe yarar. Güçlü bir şekilde ön yargılı olması beklenen karar türlerinin ön yargı giderici müdahalelerden fayda görmesi olasıdır. Örneğin, planlama yanılgısı ön yargı giderme müdahalelerini aşırı öz güvenli planlamalara karşı yetkilendirmek için yeterince sağlam bir bulgudur.

Sorun şudur ki pek çok durumda, hatanın olası istikameti önceden bilinmez. Bunlar, psikolojik ön yargıların etkisinin yargıçlar arasında değişken ve esasen öngörülemez olduğu ve sistem gürültüsüne yol açtığı bütün durumları kapsar. Bu tür durumlarda hatayı azaltmak için, tek seferde birden fazla psikolojik ön yargı yakalamaya çalışmak için daha geniş bir ağ atmamız gerekir.

Karar Gözlemcisi

Biz bu ön yargı aramasının kararın alınmasından önce ya da sonra değil, gerçek zamanda yapılmasını öneriyoruz. Elbette insanlar, ön yargıları tarafından yanlış yönlendirilirken, onları nadiren fark ederler. Bu farkındalık eksikliğinin kendisi de bilinen bir ön yargıdır: *ön yargı kör noktası.* İnsanlar başkalarındaki ön yargıları genellikle kendilerininkinden daha kolay tanırlar. Gözlemcilerin, bir ya da birden fazla tanıdık ön yargının kişinin kararlarını ya da önerilerini etkilediğine dair teşhis işaretlerini gerçek zamanda saptayacak şekilde eğitilebileceğini düşünüyoruz.

Bu sürecin nasıl işleyebileceğini zihninizde canlandırmak için, karmaşık ve sonucu önemli bir karar almaya çalışan bir grup düşünün. Karar herhangi bir türde olabilir: bir salgına ya da başka bir krize olası tepkilerine karar veren bir hükûmet, doktorların karmaşık semptomları olan bir hasta için en iyi tedaviye karar vermeye çalıştığı bir vaka konferansı, önemli bir stratejik hamleye karar veren bir yönetim kurulu. Şimdi bu grubu izleyen ve ön yargıların grubu olası en iyi karardan uzaklaştırıp uzaklaştırmadığını saptamak için bir kontrol listesi kullanan *bir karar gözlemcisi* düşünün.

Karar gözlemciliği kolay bir rol değildir ve hiç şüphesiz bazı örgütlerde gerçekçi değildir. Nihai karar alıcılar, onlarla savaşmakta kararlı olmadığı zaman, ön yargıları saptamak hiçbir işe yaramaz. Hatta karar gözlemi sürecini başlatan ve karar gözlemcisinin rolünü

destekleyenler karar alıcılar olmalıdır. Kendinizi karar gözlemcisi olarak atamanızı kesinlikle önermiyoruz. Bu şekilde ne dost kazanır ne de insanları etkileyebilirsiniz.

Ancak gayriresmî deneyler, bu yaklaşımla gerçek ilerlemenin kaydedilebileceğini öneriyor. En azından, bu yaklaşım doğru şartlar altında, özellikle bir örgütün ya da ekibin liderleri çabaya kendilerini adadığı, karar gözlemcisinin iyi seçildiği ve kendisi ciddi ön yargılara açık olmadığı durumlarda işe yarar.

Bu vakalarda karar gözlemcileri üç kategoriye ayrılır. Bazı örgütlerde, rolü bir şef üstlenebilir. Bir proje ekibi tarafından hazırlanan tekliflerin sadece ana kısmını izlemek yerine, gözetmen ayrıca tekliflerin geliştirildiği *süreci* ve ekibin dinamiklerini de yakından inceler. Bu, gözlemcinin teklifin gelişimini etkilemiş olması olası ön yargılara karşı uyanık olmasını sağlar. Başka örgütler her ekibin bir üyesini, ekibin "ön yargı avcısı" tayin edebilir; karar sürecinin bu muhafızı takım arkadaşlarına onları yanlış yönlendirebilecek ön yargıları gerçek zamanda hatırlatır. Bu yaklaşımın olumsuz yanı karar gözlemcisinin ekip içinde şeytanın avukatı konumuna yerleştirilmesi ve politik sermayesini kolayca tüketebilecek olmasıdır. Son olarak, bazı örgütler tarafsız bakış açısı avantajına (ve iç bilgi ve maliyet anlamında bazı dezavantajlara sahip) bir dış gözlemciden de yardım alabilir.

Etkili olmak için karar gözlemcilerinin bir miktar eğitime ve araçlara ihtiyacı olur. Bu araçlardan biri saptamaya çalıştıkları ön yargıların bir kontrol listesidir. Bir kontrol listesinden yardım almanın dayanağı çok açık: Kontrol listeleri yüksek riskli bağlamlarda kararları iyileştirmek konusunda uzun bir geçmişe sahipler ve geçmiş hataların tekrarlanmasını önlemeye özellikle uygunlar.

Bir örnek. Amerika Birleşik Devletleri'nde federal bürolar havayı ya da suyu temizlemek, iş yerinde ölümleri azaltmak, gıda güvenliğini artırmak, kamu sağlık krizlerine cevap vermek, sera gazı emisyonlarını azaltmak ya da iç güvenliği artırmak için ta-

sarlanmış önemli düzenlemeleri yürürlüğe koymadan önce resmî bir **mevzuat etki analizi** derlemek zorundalar. Analizin şartlarını sevimsiz bir adı (OMB Circular A-4) olan ve yaklaşık elli sayfayı bulan yoğun ve teknik bir evrak belirler. Şartlar açıkça ön yargıya karşı koymak için tasarlanmıştır. Resmî bürolar düzenlemeye neden ihtiyaç olduğunu açıklamak, daha katı ve daha gevşek alternatifleri, maliyet ve getirileri değerlendirmek, bilgiyi ön yargısız bir şekilde sunmak ve doğru bir karşılaştırma için gelecekteki değerleri uygun bir şekilde bugüne indirgemek zorundadırlar. Ancak pek çok büroda, hükûmet görevlileri bu yoğun ve teknik belgenin gereklerine riayet etmemişlerdir. (Hatta okumamış bile olabilirler.) Federal görevliler buna karşılık olarak, büroların önemli şartlarını yok saymaları ya da gözden kaçırmaları riskini azaltmak için sadece bir buçuk sayfadan oluşan basit bir kontrol listesi ürettiler.

Bir ön yargı kontrol listesinin neye benzeyebileceğini göstermek için Ek B'yi hazırladık. Bu genel kontrol listesi sadece bir örnektir; her karar gözlemcisi hem alakasını artırmak hem de benimsenmesini kolaylaştırmak için örgütün ihtiyaçlarına uyarlanmış bir liste hazırlamak isteyecektir. Önemli olan, bir kontrol listesinin bir kararı etkileyebilecek bütün ön yargıların kapsamlı listesi olmamasıdır; bu liste en sık rastlanan ve sonucu en fazla etkileyen ön yargılara odaklanmayı amaçlar.

Uygun ön yargı kontrol listeleriyle yapılan karar gözlemi, ön yargıların etkisini sınırlandırmaya yardımcı olabilir. Gayriresmî ve küçük ölçekli birtakım çabalarda cesaretlendirici sonuçlar görmekle birlikte, bu yaklaşımın etkilerinin ya da uygulamanın çeşitli yollarının artı ve eksileri hakkında sistematik bir incelemeden haberdar değiliz. Hem uygulayıcılarda hem de araştırmacılarda, karar gözlemcileri tarafından gerçek zamanlı ön yargı giderme deneyi yapılmasına ilham olmayı umuyoruz.

Gürültüyü Azaltmak: Karar Hijyeni

Ön yargı sıklıkla görebildiğimiz ve hatta açıklayabildiğimiz hatadır. Yönlüdür; bir dürtmenin ön yargının yıkıcı etkilerini sınırlayabilmesinin ya da yargıyı güçlendirme çabasının belli ön yargılarla mücadele edebilmesinin nedeni budur. Ayrıca sıklıkla görülebilir; bir gözlemcinin ön yargıları bir karar alınırken gerçek zamanda teşhis etmeyi umabilmesi bu yüzdendir.

Öte yandan gürültü kolayca göremediğimiz ya da açıklayamadığımız öngörülemez bir hatadır. Bu nedenle, çok ağır hasara neden olduğunda bile sonu sık sık yok sayarız. Bu nedenle, önleyici hijyen önlemleri tıbbi müdahaleler için neyse, gürültüyü azaltma stratejileri de ön yargıyı giderme için aynı şeydir: Amaç belirsiz bir potansiyel hata yelpazesini gerçekleşmeden önlemektir.

Gürültüyü azaltmaya bu yaklaşıma *karar hijyeni* diyoruz. Ellerinizi yıkadığınız zaman, hangi mikroptan kaçındığınızı tam olarak bilmezsiniz; sadece el yıkamanın çeşitli mikroplara karşı iyi bir önlem olduğunu bilirsiniz (özellikle salgın dönemlerinde, ama sadece o zaman değil). Benzer şekilde karar hijyeninin prensiplerine uymak, gürültüyü azaltan teknikleri altta yatan hangi hatalardan kaçınılmasına yardım ettiğinizi bilmeden benimsediğiniz anlamına gelir.

El yıkama benzetmesini özellikle seçtik. Hijyen önlemleri sıkıcı olabilir. Faydaları direkt olarak görünmez; hangi sorunun yaşanmasını önlediklerini hiçbir zaman bilemeyebilirsiniz. Tam tersine, sorunlar çıktığı zaman, hijyene riayette belli bir bozulmaya kadar iz süremeyebilirsiniz. Bu nedenlerle, el yıkama kuralına itaati sağlamak öneminin fazlasıyla farkında olan sağlık bakım çalışanları arasında bile zordur.

El yıkama ve diğer önlem biçimleri gibi, karar hijyeni de paha biçilmez ama nankördür. İyi teşhis edilmiş bir ön yargıyı düzeltmek size en azından somut bir başarma hissi verebilir. Ancak gürültüyü azaltan prosedürler için öyle olmaz. Bu prosedürler

istatistiki açıdan birçok hatayı önlerler. Ancak *hangi* hataların önlendiğini hiçbir zaman bilemezsiniz. Gürültü görünmez bir düşmandır ve görünmez bir düşmanın saldırısını önlemek sadece görünmez bir zafer getirir.

Gürültünün ne çok hasara sebep olabileceği düşünüldüğünde bu görünmez zafer, yine de mücadeleye değer. Bundan sonraki bölümler adli tıp, tahmin, tıp ve insan kaynaklarının da aralarında olduğu pek çok alanda kullanılan birçok karar hijyeni stratejisini ele alacak. Bölüm 25'te bu stratejileri gözden geçirecek ve gürültüyü azaltmaya entegre bir yaklaşımda nasıl bir araya getirilebileceklerini göstereceğiz.

Ön Yargıyı Giderme ve Karar Hijyeninden Bahsederken

"Tam olarak hangi ön yargıyla savaştığınızı ve sonucu hangi yönde etkilediğini biliyor musunuz? Bilmiyorsanız muhtemelen birden fazla ön yargı iş başındadır ve hangisinin baskın çıkacağını öngörmek zordur."

"Bu kararı tartışmaya başlamadan önce bir karar gözlemcisi atayalım."

"Karar sürecinde karar hijyenine özen gösterdik; kararın olabildiğince iyi olma şansı çok yüksek."

BÖLÜM 20

Adli Tıpta Bilgi Dizilimi

2004 yılında Madrid'de trenlere yerleştirilen bir dizi bomba 192 kişinin ölümüne, 2.000'den fazla insanın yaralanmasına neden oldu. Olay mahallinde bulunan bir naylon torbanın üstündeki bir parmak izi Interpol tarafından dünyanın her yerindeki emniyet teşkilatlarına iletildi. Günler sonra, Federal Soruşturma Bürosu'nun (FBI) suç laboratuvarı parmak izinin Oregon'da yaşayan Brandon Mayfield adında bir Amerikan vatandaşına ait olduğunu kesin olarak saptadı.

Mayfield makul bir şüpheli gibi görünüyordu. Amerikan ordusunun eski bir subayıydı ve Mısırlı bir kadınla evlendikten sonra Müslümanlığa geçmişti. Avukat olarak Taliban'a katılmak için Afganistan'a gitmeye çalışmakla suçlanan (ve daha sonra hüküm giyen) adamları temsil ediyordu. FBI'ın izleme listesindeydi.

Mayfield gözetim altına alındı, evine dinleme cihazları yerleştirildi, evi arandı, telefonları dinlendi. Bu titiz çalışmadan elle tutulur bir bilgi elde edilemeyince, FBI onu tutukladı. Ama hiçbir zaman resmî olarak suçlanmadı. Mayfield on yıldır ülkeden ayrılmamıştı. O gözaltındayken, FBI'ı daha önce Mayfield'i parmak izi için olumsuz bir eşleşme olarak gördükleri konusunda

bilgilendirmiş olan İspanyol dedektifler parmak izini başka bir şüpheliyle eşleştirdiler.

Mayfield iki hafta sonra serbest bırakıldı. Bir süre sonra Amerikan hükûmeti ondan özür diledi, 2 milyon dolarlık bir tazminat ödedi ve hatanın nedenleri konusunda kapsamlı bir soruşturma başlatıldı. Kilit bulgu şu oldu: "Hata bir metodoloji ya da teknoloji sorunu değil, bir insan hatasıydı."

Neyse ki bu tür insan hataları çok nadirdir. Buna rağmen çok eğiticidirler. Amerika Birleşik Devletleri'nin en iyi parmak izi uzmanları bir parmak izini nasıl oldu da olay mahallinin yakınından bile geçmemiş bir insana ait olarak değerlendirebildiler? Bunu öğrenmek için, önce parmak izi incelemesinin nasıl yapıldığını ve profesyonel yargının diğer örnekleriyle nasıl bir ilişkisi olduğunu anlamalıyız. Kesin bir bilim olarak düşünme eğiliminde olduğumuz adli parmak izi incelemesi aslında incelemeyi yapan kişilerin psikolojik ön yargılarına açık. Bu ön yargılar tahmin ettiğimizden daha fazla gürültü ve dolayısıyla daha fazla hata yaratabiliyorlar. Adli tıp toplumunun nasıl bütün ortamlara uygulanabilecek bir karar hijyeni stratejisini –yargıya varmak için kullanılan bilgi akışı üstünde sıkı bir kontrol– yürürlüğe koyarak bu sorunu ortadan kaldırmak için adımlar attığını göreceğiz.

Parmak İzleri

Parmak izleri parmaklarımızın sürtünme çıkıntılarının dokunduğumuz yüzeylerde bıraktığı izlerdir. Eski zamanlarda parmak uçlarının bariz teşhis işaretleri olarak kullanıldığı örnekler olmakla birlikte, modern parmak izi biliminin başlangıcı, İskoç bir hekim olan Henry Faulds'un parmak izlerinin bir teşhis tekniği olarak kullanılmasını öneren ilk bilimsel makaleyi yayımladığı on dokuzuncu yüzyılın sonuna dayanıyor.

Bunu izleyen yıllarda, parmak izleri sabıka kayıtlarında teşhis izi olarak ilgi görmeye başladı ve kademeli olarak Alphonse Bertillon adında bir Fransız polis memurunun geliştirdiği antropometrik ölçüm tekniklerinin yerini aldı. Bertillon 1912 yılında, parmak uçlarının karşılaştırılması için resmî bir sistem kodlamıştı. Daha önce kalabalıkların bilgeliğinin kâşifi olarak karşılaştığımız Sör Francis Galton da İngiltere'de benzer bir sistem geliştirmişti. (Yine de bu kurucu babaların nadiren takdir görmesine şaşırmamalı. Galton parmak izlerinin insanların ırklarına göre sınıflandırılmasında işe yarayacağına inanıyordu ve Bertillon da, muhtemelen Yahudi karşıtı ön yargıları nedeniyle, Alfred Dreyfus'un 1894 ve 1899'daki yargılanmaları sırasında bilirkişi olarak sonuç üzerinde etkili olacak şekilde –ve kusurlu– tanıklık yaptı.)

Polis memurları kısa süre sonra parmak izlerinin tekrarlı olarak suç işleyen kişilerin teşhisine yardım etmekten daha fazlasını yapabileceğini keşfettiler. 1892'de, Arjantinli bir polis memuru, Juan Vucetich bir olay yerinde bırakılmış bir parmak izini bir şüphelinin başparmağıyla karşılaştıran ilk kişi oldu. O zamandan beri, *olay yeri parmak izlerini* toplama ve *örnek izlerle* (kontrollü şartlarda bilinen bireylerden alınan izler) karşılaştırma pratiği parmak izi biliminin en kati uygulaması oldu ve adli tıbbın en yaygın kullanılan biçimini oluşturdu.

Bir elektronik parmak izi okuyucusuna denk geldiyseniz (pek çok ülkede göçmenlik bürolarının kullandığı türde) muhtemelen parmak izini dolambaçsız, mekanik ve kolayca otomatikleştirilen bir görev olarak düşünüyorsunuzdur. Ancak bir olay yerinden alınmış bir izi örnek izlerle karşılaştırmak iki temiz parmak izini eşleştirmekten çok daha hassas bir iştir. Parmaklarınızı özellikle parmak izi görüntüsünü kaydetmek için yapılmış bir makineye sıkıca bastırırsanız, temiz ve standardize bir imge yaratırsınız. Aksine, olay yeri izleri genellikle kısmi, belirsiz, bulaşık ve çarpılmış olurlar; kontrollü ve bu işe ayrılmış bir alanda alınan bir

parmak iziyle aynı nicelik ve nitelikte bilgi sağlamazlar. Olay yeri izleri sıklıkla aynı insana ya da bir başkasına ait başka izlerle üst üste biner, kir ve yüzeyde bulunan diğer insan eseri kalıntılara karışırlar. Bir şüphelinin örnek izleriyle eşleşip eşleşmediklerine karar vermek uzman yargısı gerektirir. Bu, insan parmak izi uzmanlarının işidir.

Önlerine bir olay yeri izi geldiği zaman, uzmanlar rutin olarak, ACE-V adı verilen ve analiz, karşılaştırma, değerlendirme ve teyidi içeren bir süreci kullanırlar. Önce karşılaştırma için yeterli değere sahip olup olmadığını saptamak için olay yeri izini analiz etmeleri gerekir. Eğer yeterli değer mevcutsa izi örnek bir izle karşılaştırırlar. Bu karşılaştırma bir değerlendirmeye uzanır ve değerlendirme bir *teşhis* (izler aynı insana aittir) ya da bir *dışlama* (izler aynı insana ait değildir) veya belirsiz bir kararla son bulabilir. Teşhis kararı dördüncü adımı getirir, bir başka uzmanın teyidi.

Onlarca yıldır bu sürecin güvenilirliği sorgulanmadı. Görgü tanıklarının ifadeleri tehlikeli derecede güvenilmez ve hatta itiraflar bile yanlış olabilirken, parmak izleri en güvenilir kanıt türü olarak kabul edildi (en azından DNA analizinin ortaya çıkmasına kadar). 2002 yılına kadar, Amerikan mahkemelerinde hiçbir parmak izi kanıtına başarıyla meydan okunmamıştı. Örneğin o dönemde FBI'ın internet sayfası ısrarcıydı: "Parmak izleri kişisel teşhis için *şaşmaz* bir araçtır." Hataların olduğu nadir durumlarda, suç beceriksizlik ya da sahteciliğe yüklenirdi.

Parmak izi kanıtına uzun süre kısmen hatalı olduğunu kanıtlamanın zorluğu nedeniyle meydan okunmadı. Bir parmak izinin gerçek değeri, yani suçu aslında kimin işlediğinin kesin referansı, genellikle bilinmez. Mayfield ve benzer birkaç vakada, hata özellikle korkunçtu. Ama genel olarak, eğer bir şüpheli uzmanın çıkarımlarına itiraz ederse, parmak izi kanıtı, elbette, daha güvenilir kabul edilir.

Gerçek değeri bilmemenin ne sıra dışı ne de gürültüyü ölçmeye engel olduğunu fark ettik. Parmak izi analizinde ne kadar gürültü var? Ya da daha net olmak gerekirse, parmak izi uzmanlarının, yargıçların ya da risk değerlendirme uzmanlarının aksine bir sayı üretmeseler de kesin kararlar aldıkları düşünüldüğünde, hangi sıklıkla görüş ayrılığı yaşıyorlar ve neden? University College London'da bilişsel nörobilim araştırmacısı olan Itiel Dror yola bunu incelemek için çıkan ilk kişi oldu. Gürültü sorununun olmadığı varsayılan bir alanda bir dizi gürültü denetimi gerçekleştirdi.

Parmak İzi Analizinde Durum Gürültüsü

Bir biliş bilimcinin –bir psikoloğun– parmak izi uzmanlarına meydan okuması tuhaf görünebilir. Sonuçta TV'de *CSI: Las Vegas* ve CSI'ın çeşitli diğer dizilerinde görmüş olabileceğiniz gibi, lateks eldiven giyen, mikroskop kullanan sıkı bilim insanı tiplerinden bahsediyoruz. Ancak Dror parmak izi incelemenin açıkça bir yargı meselesini olduğunu fark etti. Ve biliş konusunda çalışan bir nörobilimci olarak, yargının olduğu yerde gürültünün de olması gerektiği yönünde akıl yürüttü.

Bu hipotezi denemek için önce durum gürültüsüne; *aynı* kanıta ikinci kez bakan *aynı* uzmanların yargıları arasındaki değişkenliğe odaklandı. Dror'un ifadesiyle, "Uzmanlar kendi kendileriyle tutarlı olmak açısından güvenilir değillerse, yargılarının ve profesyonelliklerinin temeli tartışmalıdır."

Parmak izleri bir durum gürültüsü denetimi için kusursuz bir test havuzu sunuyor çünkü bir doktorun ya da yargıcın karşılaştığı vakaların aksine, iz çiftleri kolayca akılda kalmaz. Elbette uzmanların izleri hatırlamaması için aradan uygun bir zaman geçmeli. (Dror'un çalışmalarında birtakım cesur ve açık görüşlü uzmanlar, *beş yıl içinde herhangi bir zamanda,* çalışmalara bilgileri olmadan katılmayı kabul ettiler.) Ek olarak, becerilerinin

sınandığının farkında olmamaları için, deney, uzmanların rutin çalışma akışı içinde gerçekleşmeli. Bu şartlar altında, uzmanın yargısı bir testten diğerine değişiyorsa durum gürültüsü söz konusu demektir.

Adli Tıp Onay Ön Yargısı

Orijinal çalışmalarının ikisinde, Dror önemli bir husus ekledi. Uzmanların bazıları izleri ikinci kez gördüklerinde, vaka hakkında ön yargıya neden olabilecek ek bilgiye maruz bırakıldılar. Örneğin, daha önce izlerin eşleştiğini saptayan uzmanlara bu kez "şüphelinin bir tanığı olduğu" ya da "ateşli silah kanıtlarının o olmadığını düşündürdüğü" söylendi. İlk seferde bir şüphelinin masum olduğu ve izlerin tutarsız olduğu sonucuna varan diğer uzmanlara ise ikinci seferde "dedektifin şüphelinin suçlu olduğuna inandığı", "görgü tanıklarının şüpheliyi teşhis ettiği" ya da "şüphelinin suçu itiraf ettiği" söylendi." Dror bu deneyi uzmanların "ön yargılı olma ihtimalinin" test edilmesi olarak adlandırdı çünkü sağlanan bağlamsal bilgi, belirli bir istikamette psikolojik bir ön yargıyı (onay ön yargısını) harekete geçiriyordu.

Gerçekten de uzmanların ön yargıya açık oldukları anlaşıldı. Aynı uzmanlar daha önce gördükleri izleri yeniden ama bu kez ön yargıya sebep olabilecek bilgi eşliğinde değerlendirdiklerinde, yargıları değişiyordu. İlk çalışmada beş uzmandan dördü, bir dışlamayı düşündüren güçlü bağlamsal bilgi sunulduğunda, daha önceki teşhis kararlarını değiştirdiler. İkinci çalışmada ise altı uzman dört iz çiftini yeniden incelediler ve ön yargıya sebep olabilecek bilgi yirmi dört kararın dördünde değişikliğe sebep oldu. Şüpheye yer bırakmamak için, kararlarının çoğunun değişmediğini ama bu tür kararlarda altıda birlik bir değişikliğin bile büyük kabul edilebileceğini belirtelim. Bu bulgular o zamandan bu yana, başka araştırmacılar tarafından da tekrarlandı.

Öngörülebileceği gibi, uzmanlar fikirlerini değiştirmeye, ilk kararın zaten zor alındığında, ön yargıya neden olan bilginin güçlü ve değişiklik katiden kati olmayan bir karara dönüş şeklinde olduğunda daha yatkındılar. Bununla birlikte, "usta parmak izi uzmanlarının, izin içerdiği gerçek bilgi yerine bağlama göre karar alması" hayli can sıkıcıydı.

Ön yargıya sebep olan bilginin etkisi sadece uzmanın çıkarımıyla (teşhis, şüpheli bulma ya da dışlama) sınırlı değil. Ön yargıya sebep olan bilgi o algının *nasıl* yorumlandığına ek olarak uzmanın *ne* algıladığını da değiştiriyor. Dror ve çalışma arkadaşları ayrı bir çalışmada ön yargılı bir bağlama sokulan uzmanların bu tür bir bilgiye maruz kalmayan uzmanlarla aynı şeyleri görmediğini ortaya koydular. Olay yerinde bulunan ize hedef bir örnek iz eşlik ettiğinde, uzmanlar olay yeri izini tek başına gördükleri zamana göre önemli ölçüde daha az detay (küçük detaylar) gözlemliyorlardı. Daha sonra yapılan, bağımsız bir çalışma bu sonucu doğruladı ve "nasıl olduğunun belli olmadığını" ekledi.

Dror ön yargıya sebep olan bilginin etkisi için bir terim yarattı: *adli tıp onay ön yargısı*. Bu ön yargı o zamandan beri kan örüntüsü analizi, kundakçılık soruşturmaları, iskelet kalıntılarının analizi ve adli tıp patolojisi gibi diğer adli tıp tekniklerinde de gözlemleniyor. Yaygın olarak adli tıp biliminde yeni altın standart olarak kabul edilen DNA analizleri bile, en azından uzmanların karmaşık DNA karışımlarını değerlendirmeleri gerektiğinde, onay ön yargısına açık olabiliyor.

Adli tıp uzmanlarının onay ön yargısına açıklığı sadece kuramsal bir kaygı değil, çünkü aslında adli tıp uzmanlarının ön yargılı bilgiye maruz kalmaması için uygulanan sistematik bir önlem yok. Uzmanlar genellikle bu tür bilgileri onlara verilen kanıta eşlik eden ileti mektuplarından alıyorlar. Uzmanlar ayrıca genellikle polisle, savcılarla ve diğer uzmanlarla doğrudan iletişim hâlinde oluyorlar.

Onay ön yargısı bir soruna daha neden oluyor. Bir teşhisin onaylanmadan önce başka bir uzman tarafından bağımsız olarak doğrulanması ACE-V prosedürünün de bir parçası olan, hatalara karşı önemli bir güvencedir. Ancak sıklıkla sadece teşhisler bağımsız olarak doğrulanır. Doğrulayıcı uzman ilk sonucun teşhis olduğunu bildiği için, sonuç güçlü bir onay ön yargısıdır. Dolayısıyla doğrulama adımı, normalde bağımsız yargıların bir araya toplanmasından beklenen faydayı sağlamaz çünkü doğrulamalar aslında bağımsız değildir.

İki değil üç FBI uzmanının hatalı teşhise imza attığı Mayfield vakasında bir onay ön yargısı şelalesi iş başında görünüyor. Hatanın daha sonraki soruşturmasının ortaya çıkardığı gibi, ilk uzman parmak izi veri tabanlarında olası bir eşleşme arayan otomatik sistemin "korelasyon gücünden" etkilenmiş görünüyordu. Uzmanın görüntüde Mayfield'ın biyografik detaylarına maruz kalmamış olmasına rağmen, ilk aramayı gerçekleştiren bilgisayarlı sistemin sağladığı sonuçlar, "son derece önemli ve kritik bir vakada çalışıyor olmanın getirdiği yoğun baskıyla birleşerek" ilk onay ön yargısını üretmeye yetmişti. İlk uzmanın hatalı teşhisinin ardından, rapor "sonraki incelemelerin kusurlu" olduğu bilgisiyle devam ediyordu. İlk uzman çok saygı duyulan bir şef olduğu için, "bürodaki diğer çalışanların görüş ayrılığı bildirmesi gittikçe daha zorlaşmıştı." İlk hata kopyalanıp büyüyerek Mayfield'in suçluluğunun neredeyse kesinleşmesi noktasına ulaştı. Hatta mahkeme tarafından Mayfield'in savunması adına kanıtı incelemek için atanan hayli saygın bağımsız uzman bile, teşhisi doğrulamak konusunda FBI ile aynı görüşü bildirdi.

Aynı fenomen diğer adli tıp disiplinlerinde ve bunların aralarında da iş başında olabilir. Örtülü olay yeri parmak izlerinin teşhisi en nesnel adli tıp disiplinlerinden biri olarak bilinir. Parmak izi uzmanları ön yargılı olabiliyorsa, diğer alanlardaki uzmanlar da olabilirler. Dahası bir ateşli silah uzmanı parmak izlerinin eşleşti-

ğini bilirse, bu bilgi o uzmanın yargısını da taraflı kılabilir. Ve bir adli tıp odontolojisti DNA analizinin bir şüpheliyi teşhis ettiğini bilirse, bu uzmanın ısırık izinin şüphelininkine uymadığını öne sürme olasılığı muhtemelen azalır. Bu örnekler ön yargı şelaleleri vesvesesini artırıyor. Tıpkı Bölüm 8'de ele aldığımız grup kararlarında olduğu gibi, onay ön yargısının teşvik ettiği başlangıçtaki bir hata ikinci uzmanı da etkileyen, ön yargıya sebep olan bilgiye dönüşüyor ve ikinci uzmanın yargısı üçüncü uzmanı etkiliyor ve bu böyle sürüp gidiyor.

Ön yargıya sebep olan bilginin çeşitliliğe neden olduğunu kanıtladıktan sonra, Dror ve çalışma arkadaşları durum gürültüsü konusunda daha fazla kanıtı ortaya çıkardılar. Parmak izi uzmanları ön yargıya neden olan bilgilere maruz kalmadıklarında bile, bazen daha önce gördükleri izler hakkında fikir değiştirebiliyorlar. Tahmin edeceğimiz gibi, ön yargıya neden olan bilgi sağlanmadığında değişiklikler daha nadir rastlanmakla bile, yine de oluyordu. 2012 yılında FBI'ın isteğiyle yapılan bir çalışma, yetmiş iki uzmandan yaklaşık yedi ay önce değerlendirdikleri yirmi beş çift ize yeniden bakmalarını isteyerek bu bulguyu daha geniş bir ölçekte tekrarladı. Çalışma yüksek yetkinlikte uzmanlardan oluşan geniş bir grupla, parmak izi uzmanlarının bazen durum gürültüsüne açık olduklarını doğruladı. On karardan yaklaşık olarak biri değiştirilmişti. Değişikliklerin çoğu kesin olmayan kategoriye geçiş ya da o kategoriden çıkış şeklinde gerçekleşti ve hiçbiri yanlış teşhisle sonuçlanmadı. Çalışmanın en can sıkıcı iması, şüphelinin hüküm giymesine neden olan bazı parmak izi teşhislerinin bir başka zamanda "şüpheli" olarak değerlendirilebilme potansiyellerinin olmasıydı. Aynı uzmanlar aynı izlere bakarken, bağlam onları ön yargıya sevk edecek şekilde tasarlanmasa ve olabildiğince sabit tutulsa bile, kararlarında tutarsızlık olabiliyor.

Bir Miktar Gürültü ama Ne Kadar Hata?

Bu bulguların akla getirdiği pratik soru adli hataların olasılığıdır. Mahkemede tanıklık eden uzmanların güvenilirliğiyle ilgili sorular yok sayamayız: Geçerlilik güvenilirliği gerektirir çünkü en basit hâliyle, kendinizle aynı görüşte değilseniz, gerçekle aynı görüşte olmanız zordur.

Adli tıptaki hatalar tam olarak kaç hataya neden olmuştur? Yanlış hükümleri düzeltmek için çalışan kâr amacı gütmeyen bir organizasyon olan Innocence Project'in (Masumiyet Projesi) sağladığı 350 aklamanın incelenmesi, vakaların %45'inde adli tıptaki hatalı uygulamaların katkı sağlayıcı bir neden olduğunu ortaya çıkardı. Bu istatistik kulağa kötü geliyor ama yargıçlar ve jüri üyeleri için asıl önemli soru şu: Tanıklık yapmak için kürsüye çıkan uzmana ne kadar güvenmeliler ve parmak izi uzmanlarının da aralarında olduğu adli tıp uzmanlarının sonucu etkileyen hatalar yapma olasılığı nedir?

Bu sorunun en sağlam cevapları, ülkenin önde gelen bilim insanları ve mühendislerinden oluşan Başkan'ın Bilim ve Teknoloji Danışma Konseyi'nin (PCAST) 2016 yılında hazırladığı ve adli tıbbın ceza mahkemelerindeki yerini derinlemesine inceleyen bir raporda bulunabilir. Rapor parmak izi analizinin geçerliliği ve özellikle Mayfield örneğindeki gibi hatalı teşhislerin (hatalı olumlu sonuçların) olasılığı üzerine mevcut kanıtları özetliyor.

Bu kanıtlar şaşırtıcı derecede az ve PCAST'in belirttiği gibi, kanıtları üretme çalışmasının ancak son dönemde başlamış olması "can sıkıcı". En güvenilir veri, FBI'da görevli bilim insanlarının 2011 yılında bizzat yaptığı, parmak izi teşhisi doğruluğu üzerine yayımlanmış tek geniş ölçekli çalışmadan geliyor. Bu çalışma her biri yaklaşık yüz çift olay yeri ve örnek parmak izi karşılaştıran 169 uzmanı kapsıyor. Çalışmanın genel bulgusu çok az hatalı teşhis yapıldığı: Hatalı olumlu sonuç oranı altı yüzde bir kadar az.

Altı yüzde bir düşük bir hata oranı ama raporun dikkat çektiği gibi, bu oran "kamuoyunun (ve uzantısı olarak çoğu jüri üyesinin) parmak izi analizlerinin isabetliliği konusunda çok eskiye dayanan iddialara dayanarak inanmak isteyeceğinden *çok daha yüksek*." Dahası, bu çalışma ön yargıya sebep olan bağlamsal bilgi içermiyordu ve katılımcı uzmanlar bir teste girdiklerini biliyorlardı. Bu bilgi, çalışmanın gerçek vaka çalışmalarında oluşabilen hataları hafife almasına neden olmuş olabilirdi. Daha sonra Florida'da yapılan bir çalışma, daha yüksek hatalı sonuç sayısına ulaştı. Literatürdeki bulguların çeşitliliği parmak izi uzmanlarının kararlarının isabetliliği ve bu kararlara nasıl varıldığı konusunda daha fazla araştırmaya ihtiyacımız olduğunu düşündürüyor.

Öte yandan bütün çalışmalarda tutarlı görünen iç rahatlatıcı bir bulgu, uzmanların temkin yönünde hata yapma eğilimde olduğunu gösteriyor. İsabetlilik oranları kusursuz olmayabilir ama yargılarının sonuçlarının bilincindeler ve olası hataların asimetrik maliyetini hesaba katıyorlar. Parmak izinin yüksek güvenilirliği nedeniyle, hatalı bir eşleştirme trajik sonuçlar yaratabilir. Diğer hata türlerinin sonuca etkisi daha az. Örneğin, FBI uzmanlarının ifadesiyle, "çoğu vaka çalışmasında, bir dışlama, kesin olmayan, şüpheli bir sonuçla aynı işlemsel anlamı taşıyor." Başka bir deyişle, bir cinayet silahında bir parmak izinin bulunması hüküm vermek için yeterli ama o izin yokluğu, şüpheliyi aklamaya yetmiyor.

Parmak izi uzmanlarının temkiniyle ilgili gözlemimizle tutarlı olarak, kanıtlar uzmanların bir teşhis kararına varmadan önce iki kez –hatta iki kezden fazla– düşündüklerini gösteriyor. FBI'ın teşhislerin isabetliliği konulu çalışmasında, "eşleşen" çiftlerin (yani olay yerindeki izle örnek izin aynı insana ait olduğu durumların) üçte birinden azı (doğru) teşhis olarak kabul edildi. Uzmanlar ayrıca yanlış olumsuz dışlamadan çok daha az hatalı olumlu teşhis kararı veriyorlar. Ön yargıya açıklar ama bu, iki istikamette aynı derecede değil. Dror'un belirttiği gibi, "Adli tıp uzmanlarında 'ke-

sin olmayan şüpheli' sonuca yönlendirecek ön yargıyı yaratmak, kesin 'teşhis' sonucuna yönlendirmekten daha kolay."

Uzmanlar hatalı teşhisi ne pahasına olursa olsun kaçınılması gereken ölümcül bir günah olarak kabul edecek şekilde eğitiliyorlar. Haklarını vermek gerekir ki bu prensibe uygun hareket ediyorlar. Özen seviyelerinin Mayfield vakasındaki ve diğer birkaç önemli vakadaki yanlış teşhislerin çok nadir yaşanmasına yeteceğini ummaktan başka çaremiz yok.

Gürültüyü Dinlemek

Adli tıpta gürültü olduğu gözlemi adli tıp uzmanlarına bir eleştiri olarak görülmemelidir. Bu sadece defalarca tekrarladığımız bir gözlemin sonucudur: Yargının olduğu yerde, gürültü vardır, sandığınızdan daha fazla. Parmak izlerinin analizi objektif bir görev gibi görünebilir; hatta bu yüzden çoğumuz ilk anda onu bir yargı biçimi olarak kabul etmeyiz. Yine de tutarsızlığa, anlaşmazlığa ve zaman zaman hataya yer olan bir görevdir. Parmak izi teşhislerindeki hata oranı ne kadar düşük olursa olsun, sıfır değildir ve PCAST'in raporladığı gibi, jüriler bundan haberdar edilmelidir.

Gürültüyü azaltmanın ilk adımı, elbette, olma olasılığını kabul etmektir. Bu kabulleniş, parmak izi toplumunun, çoğu Dror'un gürültü denetimi konusunda başlangıçta hayli şüpheci olan üyeleri için çok kolay bir şey değildir. Bir uzmanın vakayla ilgili bilgiden farkında olmadan etkilenebilmesi birçok uzmanın canını sıktı. Dror'un çalışmasına cevap olarak, Parmak İzi Topluluğu yönetimi "Bir uzman karar alma sürecinde... her iki yöne yalpalayabilmesi... ancak Disneyland'da çalışacak kadar toy olduğu anlamına gelir", ifadesini kullandı. Önemli bir adli tıp laboratuvarının yöneticisi vaka bilgisine –özellikle uzmanda ön yargıya sebep olabilecek türde bilgiye– erişmenin "[uzmanlara] işlerinden *yargılarını değiştirmeden* keyif alma imkânı veren kişisel tatmin sağladığını"

belirtti. FBI bile Mayfield'la ilgili iç soruşturmasında, "olay yeri parmak izi inceleme uzmanlarının, rutin olarak, önceki uzmanların sonuçlarını bildikleri doğrulama çalışmaları yaptıklarını *ve buna rağmen o sonuçların uzmanın çıkarımlarını etkilemediğini*" belirtti. Bütün bu yorumlardan onay ön yargısının varlığının inkâr edildiği sonucu çıkıyor.

Adli tıp uzmanları ön yargı riskinin farkında olduklarında bile, ön yargı kör noktasına, yani ön yargının varlığını kendilerinde değil, başkalarında kabul etme eğilimine bağışık değiller. Yirmi bir ülkede dört yüz profesyonel adli tıp uzmanı arasında yapılan bir ankette katılımcıların %71'i "bilişsel ön yargının adli tıbbın bütününde endişe sebebi olduğu" konusunda hemfikirken, "kendi yargısının da bilişsel ön yargıdan etkilendiğini" düşünenlerin oranı sadece %26 idi. Başka bir deyişle, bu adli tıp profesyonellerinin yarıya yakını meslektaşlarının yargılarının gürültülü olduğuna ama kendilerininkinin öyle olmadığına inanıyorlar. Gürültü, işi görünmez olanı görmek olan insanlar için bile, görünmez bir problem olabilir.

Bilgiyi Sıralamak

Dror ve çalışma arkadaşlarının ısrarı sayesinde, tavırlar değişiyor ve analizlerindeki hataları azaltmak için yeni önlemler alan adli tıp laboratuvarlarının sayısı her geçen gün artıyor. Örneğin PCAST raporu onay ön yargısını en az seviyeye indirmek için FBI laboratuvarının prosedürlerini yeniden düzenlemesini önerdi.

Atılması gereken sistemsel adımlar nispeten basit. Pek çok alanda uygulanabilirliği olan bir karar hijyeni stratejisine karşılık geliyorlar: *henüz olgunlaşmamış sezgilerin vaktinden önce oluşmasını sınırlandırmak için bilginin sıralanması.* Her yargıda bazı bilgiler önemli ve alakalıdır, bazıları değildir. Daha fazla bilgi her zaman daha iyi değildir; özellikle yargıcı vaktinden önce, henüz

olgunlaşmamış bir sezgi oluşturmaya yönlendirerek ön yargıya sebep olma potansiyeli varsa.

Bu açıdan, adli tıp laboratuvarlarında benimsenen yeni prosedürler, uzmanlara sadece ihtiyaçları olan bilgiyi ve ihtiyaç duydukları zaman sağlayarak yargılarının bağımsızlığını korumayı amaçlıyor. Başka bir deyişle, laboratuvar onları olabildiğince bilgisiz bırakıyor ve bilgiyi sadece kademeli olarak ortaya sürüyor. Dror ve çalışma arkadaşlarının bunu yapmak için düzenlediği yaklaşımın adı *doğrusal sıralı ifşa*dır.

Dror'un aynı karar hijyeni stratejisini temsil eden bir önerisi daha var: Uzmanlar yargılarını her adımda belgelemelidir. Olay yerindeki bir parmak izinin analizini, eşleşip eşleşmediğine karar vermek için ellerindeki örnek parmak izlerine bakmadan önce belgelemelidirler. Bu adım sıralaması uzmanların sadece aradıkları şeyi görme riskinden sakınmasına yardımcı olur. Ayrıca uzmanlar kanıt hakkındaki yargılarını ön yargıya sebep olma riski olan bağlamsal bilgiye erişim kazanmadan kaydetmelidirler. Bağlamsal bilgiyle karşı karşıya geldikten sonra fikirlerini değiştirirlerse, bu değişiklikler ve değişikliklerin gerekçeleri belgelenmelidir. Bu şart başlangıçtaki bir sezginin bütün sürece ön yargı katma riskini sınırlandırır.

Aynı mantık karar hijyeninin önemli bir kısmını oluşturan üçüncü bir öneriye ilham olmuştur. İlk kişinin yaptığı bir teşhisi doğrulamak için ikinci bir uzman çağrıldığında, ikinci uzman birinci uzmanın yargısının farkında olmamalıdır.

Adli tıpta gürültünün varlığı, ölüm-kalım seviyesinde sonuçlar üretme potansiyeli nedeniyle elbette endişe sebebidir. Ama aynı zamanda aydınlatıcıdır. Parmak izi teşhisinde hata olasılığından bu kadar uzun süre tamamen bihaber kalmamız, uzman insan yargısına olan güvenimizin bazen nasıl abartılabildiğini ve bir gürültü denetiminin beklenmedik bir gürültü miktarını açığa çıkarabileceğini göstermektedir. Bu eksikleri nispeten basit süreç

değişiklikleriyle hafifletme becerisi, kararların kalitesini iyileştirmeyi önemseyen herkes için cesaret verici olmalı.

Bu örneğin işaret ettiği başlıca hijyen stratejisi –bilgiyi sıralama– durum gürültüsüne karşı bir koruma olarak çok geniş bir uygulanabilirliğe sahip. Belirttiğimiz gibi, durum gürültüsü, ruh hâli ve dışarıdaki hava sıcaklığının da aralarında olduğu sayısız tetikleyici tarafından yönlendirilir. Bütün bu tetikleyicileri kontrol etmeyi umamazsınız ama yargıları en bariz olanlardan korumaya çaba harcayabilirsiniz. Örneğin, yargıların öfke, korku ya da diğer duygularla değiştirilebileceğini biliyorsunuz ve belki de mümkünse zamanda, durum gürültüsü tetikleyicilerinin farklı olabileceği, farklı noktalarda yargılarınıza geri dönmenin iyi bir uygulama olduğunu fark etmişsinizdir.

Yargınızın durum gürültüsünün bir başka tetikleyicisi olan bilgi tarafından –doğru bilgi bile olsa– değiştirilmesi olasılığı ise daha az barizdir. Parmak izi uzmanlarının örneğinde olduğu gibi, başkalarının ne düşündüğünü bildiğiniz anda, onay ön yargısı sizi vaktinden çok önce bir izlenim oluşturmaya ve çelişen bilgileri yok saymaya yönlendirebilir. İki Hitchcock filminin isimleri durumu çok iyi özetliyor: İyi bir karar alıcı, "çok (şey) bilen adam" olmamak için, her zaman "şüphenin gölgesi"ni korumayı amaç edinmelidir.

Bilgiyi Sıralamaktan Bahsederken

"Yargının olduğu yerde gürültü vardır ve buna parmak izlerinin okunması da dâhildir."

"Bu vaka hakkında daha fazla bilgi sahibiyiz ama ön yargıya yol açmamak için, uzmanlara bildiğimiz her şeyi, onlar kararlarını vermeden önce söylemeyelim. Hatta onlara sadece kesinlikle bilmeleri gerekenleri söyleyelim."

"İkinci görüşün sahibi, ilk görüşten haberdarsa, görüşü bağımsız olmaz. Üçüncü görüşte bu daha da artar; bir ön yargı şelalesi yaşanabilir."

"Gürültüyle savaşmak için, önce varlığını kabul etmeliler."

BÖLÜM 21

Tahminde Seçme ve Bir Araya Toplama

Pek çok yargı, tahmin içerir. Bir sonraki çeyrekte işsizlik oranının kaç olması olasıdır? Önümüzdeki yıl kaç elektrikli araba satılacak? 2050'de iklim değişikliğinin etkileri ne olacak? Yeni bir bina yapmak ne kadar zaman alacak? Belli bir şirketin yıllık kazancı ne olacak? Yeni bir çalışan nasıl bir performans sergileyecek? Yeni bir hava kirliliği düzenlemesinin maliyeti ne olacak? Seçimi kim kazanacak? Bu tür soruların cevaplarının önemli sonuçları olur. Özel kurumların yanı sıra kamu kurumlarının da temel tercihleri genellikle bu cevaplara bağlıdır.

Tahmin analizcileri –ne zaman ve neden sorun çıktığını analiz eden kişiler– ön yargı ve gürültü (tutarsızlık ve güvenilmezlik de denir) arasında keskin bir ayrım yaparlar. Tahmincilerin bazı bağlamlarda ön yargılı olduğu konusunda herkes hemfikirdir. Örneğin resmî bürolar bütçe öngörülerinde gerçek dışı bir iyimserlik sergilerler. Ortalamada, gerçekçi olmaya uzak yüksek bir ekonomik büyüme ve gerçekçi olmaya uzak düşük zararlar beklerler. Pratikteki amaçlar açısından, bu gerçek dışı iyimserliğin bilişsel

ön yargının mı yoksa politik değerlendirmelerin mi ürünü olduğu çok önemli değildir.

Buna ek olarak, tahminciler ayrıca aşırı öz güvenli olma eğilimindedirler: Tahminlerini nokta tahminler yerine, güven aralıklarıyla formüle etmeleri istendiğinde, olması gerekenden daha dar aralıklar seçme eğiliminde olurlar. Örneğin, devam etmekte olan üç aylık bir anket ABD şirketlerinin CFO'larından bir sonraki yıl için S&P 500 endeksinin yıllık geri dönüşünü tahmin etmelerini ister. CFO'lar iki sayı verirler: gerçek geri dönüşün altında olma ihtimalinin onda bir olduğunu düşündükleri bir minimum ve gerçek geri dönüşün aşma ihtimalinin onda bir olduğu bir maksimum. Dolayısıyla iki sayı %80'lik bir güven aralığının sınırlarıdır. Bununla birlikte, gerçekleşen geri dönüş sadece %36 oranında o aralığa denk gelir. CFO'lar tahminlerinin netliğinde fazla öz güvenlidirler.

Tahminciler aynı zamanda gürültülüdür. J. Scott Armstrong'un *Principles of Forecasting* adlı kitabı, uzmanların arasında bile, "güvensizliğin yargısal öngörüde bir hata kaynağı olduğuna" dikkat çeker. Durum gürültüsü yaygındır; tahminciler her zaman kendileriyle aynı fikirde olmazlar. Kişiler arası gürültü de yaygındır; tahminciler uzman bile olsalar, birbirleriyle de anlaşmazlığa düşerler. Hukuk profesörlerinden Yüksek Mahkeme'nin kararlarını tahmin etmelerini isterseniz, bol miktarda gürültüyle karşılaşırsınız. Uzmanlardan hava kirliliği düzenlemesinin yıllık getirilerini öngörmelerini isterseniz, örneğin üç milyar dolarla dokuz milyar dolar aralığında, çok büyük bir değişkenlikle karşılaşırsınız. Bir grup ekonomistten işsizlik ve büyüme tahminleri isterseniz yine çok büyük bir çeşitlilikle karşı karşıya kalırsınız. Daha önce gürültülü öngörülerin birçok örneğini görmüştük ve tahmin konusunda yapılan araştırmalar yenilerini açığa çıkarıyor.

Tahminleri İyileştirmek

Araştırma ayrıca gürültü ve ön yargıyı azaltmak için önerilerde bulunuyor. Bu önerileri burada kapsamlı olarak incelemeyeceğiz ama geniş bir uygulanabilirliği olan iki gürültüyü azaltma stratejisine odaklanacağız. Bu stratejilerden ilki, Bölüm 18'de değindiğimiz prensibin uygulanması: Daha iyi yargıçlar seçmek, daha iyi yargılar üretir. Diğeri ise hijyen stratejilerinin evrensel olarak en uygulanabilir olanlarından biri: çok sayıda bağımsız tahmini bir araya toplamak.

Çok sayıda tahmini bir araya toplamanın en kolay yolu, ortalamalarını almaktır. Ortalama almak matematiksel olarak gürültüyü azaltmayı garantiler; gürültüyü ortalaması alınan yargıların sayısının kareköküne böler. Bu, yüz yargının ortalamasını alırsanız gürültüyü %90 oranında azaltacağınız anlamına gelir. Dört yüz yargının ortalamasını alırsanız, %95 oranında azaltır, neredeyse tamamen ortadan kaldırmış olursunuz. Bu istatistik kuralı Bölüm 7'de bahsettiğimiz kalabalıkların bilgeliği yaklaşımının lokomotifidir.

Ortalama almak ön yargıyı azaltmak konusunda hiçbir işe yaramadığı için, toplam hata (OKH) üzerindeki etkisi, hatada ön yargı ve gürültünün oranlarına bağlıdır. Kalabalıkların bilgeliğinin, en çok yargılar bağımsız olduğunda –dolayısıyla ortak ön yargı barındırması olasılığı azaldığında– işe yaraması bu yüzdendir. Deneysel olarak çok sayıda kanıt, birden fazla tahminin ortalamasını almanın isabetliliği büyük ölçüde artırdığını düşündürmektedir. Örneğin, ekonomi tahmincilerinin borsa analizcileri için "ortak" tahmininde. Satış tahminleri, hava durumu tahminleri ve ekonomi tahminleri söz konusu olduğunda, bir grup tahmincinin ağırlıksız ortalaması çoğu bireysel tahminciden ve hatta bazen hepsinden daha iyi bir performans sergiler. Farklı yöntemlerle elde edilmiş tahminlerin ortalamasını almak da aynı etkiyi yaratır:

Farklı alanlarda otuz deneysel karşılaştırmanın analizinde, birleştirilmiş tahminler hataları %12,5'lik bir ortalamayla azaltmıştır.

Düz ortalama almak tahminleri birleştirmenin tek yolu değildir. Son yargılarının isabetliliğine göre en iyi yargıçları seçen ve az sayıdaki yargıcın (örneğin beş) yargılarının ortalamasını alan *seçme kalabalık* stratejisi de düz ortalama almak kadar etkili olabilir. Ayrıca, uzmanlığa saygı duyan karar alıcıların sadece bir araya toplamaya değil, aynı zamanda seçmeye de dayanan bir stratejiyi anlaması ve benimsemesi daha kolaydır.

Toplama tahminler üretmenin bir diğer yöntemi de bireylerin olası sonuçlara bahis oynadığı ve dolayısıyla doğru tahminlerde bulunmaya teşvik edildiği *tahmin piyasalarını* kullanmaktır. Çoğu zaman, tahmin piyasalarının çok iyi iş çıkardığı görülmüştür: Tahmin piyasasında sözleşme fiyatları bir olayın gerçekleşme olasılığının, diyelim, %70 olduğunu düşündürüyorsa, o olay yüz seferden yetmişinde gerçekleşecektir gibi. Çeşitli sektörlerden birçok şirket çeşitli görüşleri bir araya toplamak için tahmin piyasalarından faydalanır.

Çeşitli görüş açılarını bir araya toplamanın bir diğer biçimsel süreci de Delphi yöntemi olarak bilinir. Klasik hâlinde bu yöntem katılımcıların birbirlerine anonim kalarak bir moderatöre tahminde bulunduğu (ya da oyladığı) birden fazla turu içerir. Her yeni turda katılıcımlar tahminleri için nedenler sunar ve başkalarının verdiği nedenlere yine anonim olarak cevap verirler. Süreç tahminlerin birbirine yaklaşmasını teşvik eder (hatta bazen yeni yargıların, bir önceki turun yargılarının dağılımında belli bir aralığa düşmesini şart koşarak onları buna zorlar). Bu yöntem hem bir araya toplamadan, hem sosyal öğrenmeden faydalanır.

Delphi yöntemi birçok durumda çok işe yaramıştır ama yerleştirilmesi zor olabilir. Daha basit bir versiyonu olan *mini-Delphi* tek bir toplantı içinde uygulanabilir. Ayrıca *tahmin-konuşma-tahmin* adı da verilen bu metot katılımcıların önce ayrı (ve sessiz olarak)

tahminler üretmesini, sonra bu tahminleri açıklayıp gerekçelendirmesini ve son olarak tahminlere ve diğerlerinin açıklamalarına cevap olarak yeni bir tahminde bulunmasını ister. Uzlaşılan yargı o ikinci turda elde edilen bireysel tahminlerin ortalamasıdır.

İyi Yargı Projesi

Tahminin kalitesi üstüne, şu ana kadar keşfettiklerimizin çok ötesine geçen en yenilikçi çalışmaların bazıları, 2011 yılında, öne çıkan üç davranış bilimcinin İyi Yargı Projesi'ni kurmasıyla başladı. Philip Tetlock (Bölüm 11'de politik olayların uzun vadeli öngörüleriyle ilgili değerlendirmesini tartışırken karşılaşmıştık), eşi Barbara Mellers ve Don Moore tahmin konusunu ve özellikle bazı insanların bu konuda neden daha iyi olduğunu daha iyi anlamamız için bir araya geldiler.

İyi Yargı Projesi uzman değil, hayatın farklı kollarından sıradan insanlar olan on binlerce gönüllünün görevlendirilmesiyle başladı. Bu kişilere şuna benzer yüzlerce soru soruldu:

- ❑ *Kuzey Kore bu yıl sona ermeden nükleer bir silah patlatacak mı?*
- ❑ *Rusya önümüzdeki üç ay içinde bir miktar Ukrayna toprağını daha ilhak edecek mi?*
- ❑ *Hindistan ve Brezilya önümüzdeki iki yıl içinde BM Güvenlik Konseyi'nin kalıcı üyesi olacaklar mı?*
- ❑ *Önümüzdeki yıl, Avro Bölgesi'nden çekilen ülke olacak mı?*

Bu örneklerin gösterdiği gibi, proje dünya olayları hakkında büyük sorulara odaklandı. En önemlisi, bu tür soruları cevaplama çabası daha dünyevi tahminlerle aynı sorunların çoğunu ortaya atar. Eğer bir avukata bir müvekkili mahkemede kazanıp kazan-

mayacağını soruyorsa ya da bir televizyon stüdyosuna önerilen bir dizinin büyük başarı kazanıp kazanmayacağı soruluyorsa, işin içinde tahmin becerileri vardır. Tetlock ve çalışma arkadaşları bazı insanların özellikle iyi tahminciler olup olmadığını öğrenmek istiyorlardı. Ayrıca tahmin becerisinin öğretilebilir ya da en azından iyileştirilebilir bir şey olup olmadığını öğrenmek istiyorlardı.

Temel bulguları anlamak için, Tetlock ve ekibinin tahmincileri değerlendirmek için benimsediği yöntemin birtakım kilit yönlerini açıklamalıyız. Öncelikle, şansın başarıdan ya da başarısızlıktan sorumlu olacağı bir ya da birkaç tane değil, çok büyük miktarda tahmin kullandılar. Tuttuğunuz takımın bir sonraki maçında kazanacağını tahmin ederseniz ve öyle olursa bu illa iyi bir tahminci olduğunuz anlamına gelmez. Belki *her defasında* tuttuğunuz takımın kazanacağınız tahmin ediyorsunuz; stratejiniz buysa ve eğer maçların sadece yarısını kazanıyorlarsa, tahmin beceriniz özellikle etkileyici değildir. Şansın rolünü azaltmak için, araştırmacılar katılımcıların çok sayıda tahminde, ortalamada ne kadar iyi iş çıkardıklarını incelediler.

İkinci olarak, araştırmacılar katılımcılardan, tahminlerini "olacak" ya da "olmayacak" şeklinde ikili bir cevap vermek yerine bir olayın olma olasılığını belirterek yapmalarını istediler. Birçok insan için tahmin yapmak ilk seçenekteki gibi, iki duruştan birini sergilemektir. Ancak gelecekte olacak olaylar konusunda objektif bilgisizliğimiz göz önüne alındığında, olasılıkçı tahminler oluşturmak çok daha iyidir. 2016'da biri "Hillary Clinton'ın başkan seçilme ihtimali %70 deseydi" illa kötü bir tahminci olmazdı. Doğru bir şekilde, gerçekleşme olasılığının %70 olduğu söylenen şeyler, zamanın %30'unda gerçekleşmez. Tahmincilerin iyi olup olmadığını bilmek için olasılık tahminlerinin gerçeklikle örtüşüp örtüşmediğini sormalıyız. Margaret adında bir tahmincinin 500 farklı olayın olma olasılığının %60 olduğunu söylediğini varsayalım. Bu olaylardan 300'ü gerçekten olursa, o zaman Margaret'in güve-

ninin iyi *ayarlı* olduğu sonucuna varabiliriz. İyi ayar, iyi tahmin için gerekli şartlardan biridir.

Üçüncü ve ek bir ince ayar olarak, Tetlock ve ekip arkadaşları tahmincilerinden sadece bir olayın, mesela on iki ay içinde gerçekleşip gerçekleşmeyeceği konusunda *tek* bir olasılık tahmini yapmalarını istemediler. Katılımcılara yeni bilgilerin ışığında tahminlerini sürekli olarak düzeltme imkânı tanıdılar. 2016'da, İngiltere'nin 2019 yılı bitmeden Avrupa Birliği'nden ayrılma ihtimalinin sadece %30 olduğunu tahmin ettiğinizi varsayalım. "Çıkılmalı" yönündeki oyların hız kazandığını düşündüren yeni anketler çıktıkça, muhtemelen tahmininizi yukarı doğru yenilerdiniz. Referandum sonuçları öğrenildiğinde, İngiltere'nin birlikten o zaman çerçevesinde çıkıp çıkmayacağı hâlâ belirsizdi ama çok daha olası görünüyordu. (Brexit teknik olarak 2020'de gerçekleşti.)

Tetlock ve meslektaşları her yeni bilgiyle tahmincilerin tahminlerini güncellemelerine izin verdiler. Puanlama açısından, bu güncellemelerin her biri yeni bir tahmin muamelesi gördü. Bu şekilde İyi Yargı Projesi'ndeki katılımcılar haberleri takip etmeye ve tahminlerini sürekli olarak güncellemeye teşvik edildiler. Bu yaklaşım iş hayatında ve hükûmette yer alan ve tahminlerini fikirlerini değiştirdikleri için eleştirilme riskine rağmen, yeni bilgilere göre sık sık güncellemesi gereken tahmincilerden beklenen tavrın bir aynasıdır. (Bu eleştiriye çok bilinen ve bazen John Maynard Keynes'e atfedilen cevap şöyledir: "Olgular değişince, fikrimi değiştiriyorum. Peki *sen* ne yapıyorsun?")

İyi Yargı Projesi dördüncü olarak, tahmincilerin performanslarını puanlamak için Glenn W. Brier tarafından 1950'de geliştirilen bir sistem kullandı. *Brier* skorları insanların tahminleriyle aslında olanlar arasındaki farkı ölçer.

Brier skorları olasılıkçı tahminlerle bağlantılı yaygın bir sorunu aşmanın akıllıca bir yoludur: Tahmincilerin bahislerini asla cesur bir duruş takınmayarak korumaya almasını teşvik eder. Beş yüz

olayı %60 olası olarak tanımladığı ve o olayların üç yüzü gerçekleştiği için iyi ayarlı bir tahminci olarak tanımladığımız Margaret'i düşünün. Bu sonuç göründüğü kadar etkileyici olmayabilir. Margaret *her zaman* %60 yağmur ihtimali öngören bir tahminci olsa ve beş yüz günden üç yüzünde yağmur yağsa Margaret'in tahminleri iyi ayarlı ama bir pratikte işe yaramaz olurdu. Margaret özünde size her gün bir şemsiye taşımanızın fena olmayacağını söylüyor. Onu yağmurun yağdığı üç yüz günde %100 yağmur ihtimali öngören ve iki yüz kuru günde yağmura %0 ihtimal tanıyan Nicholas'la karşılaştıralım. Nicholas da Margaret gibi kusursuz bir ayara sahip; her iki tahminci de günlerin %X'inin yağmurlu olacağını öngördüklerinde, yağmur tam olarak %X'inde yağıyor. Ama Nicholas'ın tahminleri çok daha değerli: Nicholas lafı dolandırarak tahminlerini sağlama almak yerine, size bir şemsiye almanıza gerek olup olmadığını söylüyor. Teknik olarak, Nicholas'ın iyi ayara ek olarak yüksek *çözünürlüğü* olduğu söylenebilir.

Brier skorları hem iyi ayarı hem de iyi çözünürlüğü ödüllendirir. İyi bir skor üretmek için, sadece ortalamada doğru (yani iyi ayarlı) olmanız yetmez, aynı zamanda bir duruş takınmanız ve tahminler arasında fark yaratmanız (yani yüksek çözünürlük sahibi olmanız) gerekir. Brier skorları ortalama karesel hata mantığına dayanır; daha düşük skorlar daha iyidir, 0'lık bir skor kusursuz olur.

Şimdi artık nasıl puanlandıklarını öğrendiğimize göre, İyi Yargı Projesi gönüllüleri nasıl bir iş çıkardılar? Başlıca ve en önemli bulgulardan biri, gönüllülerin sarsıcı bir çoğunluğunun kötü iş çıkardığı ama öngörülerinin şanstan çok daha iyi olduğuydu. Dikkat çekici bir şekilde, bir hükûmet görevlisi grubun "yakalanan radyo sinyallerini ve diğer gizli verileri okuyabilen ortalama bir istihbarat topluluğu analizcisinden daha iyi iş çıkardığı" yorumunu yaptı. Bu karşılaştırma duraksamayı hak ediyor. İstihbarat topluluğu analizcileri isabetli tahminler yapmak için eğitilirler;

amatör değildirler. Ek olarak, gizli bilgilere erişebilirler. Ve buna rağmen süper tahminciler kadar iyi iş çıkarmıyorlar.

Sürekli Beta

Süper tahmincileri bu kadar iyi yapan nedir? Bölüm 18'deki argümanımızla tutarlı olarak, olağanüstü zeki oldukları spekülasyonunu yapabiliriz. Bu spekülasyon yanlış olmaz. GZY testlerinde süper tahminciler İyi Yargı Projesi'ndeki ortalama bir gönüllüden (ve ortalama bir gönüllü ulusal ortalamanın hayli üstünde) daha iyi iş çıkarıyorlar. Ama fark o kadar büyük değil ve zekâ testlerinde iyi sonuç alan birçok gönüllü süper tahminci olarak adlandırılmaya hak kazanamıyor. Genel zekânın dışında, makul bir şekilde, süper tahmincilerin sayılarla olağanüstü iyi olmalarını bekleyebiliriz. Ve öyledirler. Ama asıl avantajları matematik becerileri değildir; analitik ve olasılıkçı düşünmedeki rahatlıklarıdır.

Süper tahmincilerin sorunları yapılandırma ve parçalara ayırma istekliliğini ve becerisini düşünün. Büyük bir jeopolitik soru hakkında (bir ülke Avrupa Birliği'nden çıkacak mı, belli bir yerde savaş patlak verecek mi ve bir kamu görevlisine suikast yapılacak mı gibi) bütüncül bir yargı oluşturmak yerine, onu ögelere ayırırlar. "Cevabın evet olması için ne gerekli? Cevabın hayır olması için ne gerekli?" diye sorarlar. Bir iç ses ya da küresel bir önsezi sunmak yerine, çeşitli tamamlayıcı sorular sorup cevap bulmaya çalışırlar.

Süper tahminciler ayrıca dış bakış takınmak konusunda da ustadırlar ve baz oranları çok önemserler. Bölüm 13'te Gambardi probleminde açıklandığı gibi, Gambardi'nin profilindeki özelliklere odaklanmadan önce, ortalama bir CEO'nun iki yıl içinde işten kovulma ya da istifa etme olasılığını bilmenin büyük faydası olur. Süper tahminciler sistematik olarak baz oranları ararlar. Önümüzdeki yıl Çin ve Vietnam arasında sınır anlaşmazlığı nedeniyle silahlı bir

çatışma yaşanıp yaşanmayacağı sorulduğunda, süper tahminciler sadece ve hemen Çin ve Vietnam'ın şu anda nasıl geçindiğine odaklanmaz. Okudukları haber ve analizlerin ışığında bu konuda bir sezgileri olabilir. Ama tek bir olay hakkındaki önsezilerinin genel olarak iyi bir rehber olmadığını bilirler. Bunun yerine bir baz oran aramaya başlarlar. Geçmişteki sınır anlaşmazlıklarının hangi sıklıkla silahlı çatışmaya döndüğünü sorarlar. Çatışmalar nadirse, süper tahminciler işe bu olguyu dâhil ederek başlar ancak bundan sonra Çin-Vietnam durumunun detaylarına dönerler.

Kısaca, süper tahmincileri diğerlerinden ayıran saf zekâları değil, onu nasıl kullandıklarıdır. Uygulamaya koydukları beceriler Bölüm 18'de tarif ettiğimiz türde özellikle yüksek bir "aktif açık görüşlülük" seviyesi barındıran ve daha iyi yargılar üretmesi olası bir bilişsel tarzı yansıtır. Aktif açık görüşlülük testini hatırlayın: "İnsanlar inançlarına ters düşen kanıtları da hesaba katmalıdır" ve "Sizinle hemfikir olan insanlardan çok, olmayanlara dikkat etmek daha faydalıdır" gibi beyanları içerir. Besbelli, bu testte yüksek puana ulaşan insanlar yeni bilgi erişilebilir olduğu zaman, yargılarını (aşırı tepki vermeden) güncellemekten çekinmezler.

Tetlock süper tahmincilerin düşünme tarzını tarif etmek için bilgisayarcıların nihai versiyonu bir türlü piyasaya sürülmeyen ama sürekli olarak kullanılan, analiz edilen ve iyileştirilen programlar için kullandığı "sürekli beta" terimini kullanıyor. "Süper tahminciler rütbesine yükselmenin en güçlü kestiricisinin sürekli beta, kişinin kendini inançlarını güncelleme ve kendini geliştirmeye adama derecesi" olduğunu düşünüyor. Tetlock'un ifadesiyle, "Onları bu kadar iyi kılan ne olduklarından çok ne yaptıklarıdır: araştırma çabası, dikkatli düşünce ve öz eleştiri, diğer bakış açılarının toplanması ve sentezlenmesi, ayrıntılı yargı ve amansız güncelleme." Süper tahminciler bir düşünce çemberini özellikle sever: "Dene, başarısız ol, analiz et, düzelt, tekrar dene."

Tahminde Gürültü ve Ön Yargı

Bu noktada insanların süper tahminci olmak ya da en azından onlara benzer bir performans sergileyebilmek için eğitilebileceğini düşünmek cazip gelebilir. Ve hatta Tetlock ve çalışma arkadaşları da tam olarak bunu yapmaya çalıştılar. Çabaları süper tahmincilerin neden bu kadar iyi performans sergilediğini ve daha iyi performans sergilemelerinin nasıl sağlanabileceğini anlamanın ikinci aşaması olarak görülmelidir.

Tetlock ve ekibi önemli bir çalışmada sıradan (süper olmayan) tahmincileri farklı müdahalelerin daha sonraki yargıların kalitesi üzerindeki etkisini test etmek için rastgele üç gruba ayırdılar.

1. *Eğitim:* Birçok tahminci onlara olasılıkçı akıl yürütmeyi öğreterek becerilerini iyileştirmek için tasarlanmış bir eğitim aldılar. Bu eğitimde, tahminciler çeşitli ön yargıları (aralarında baz oranı yok saymak, aşırı öz güven ve onay ön yargısı da vardı), çeşitli kaynaklardan çok sayıda öngörünün ortalamasını almanın önemini ve referans sınıflarını hesaba katmayı öğrendiler.

2. *Takım olma (bir toplama biçimi):* Bazı tahmincilerden birbirlerinin tahminlerini görebildikleri ve tartışabildikleri takımlar hâlinde çalışmaları istendi. Takım olma tahmincileri karşı argümanlarla uğraşmaya ve aktif olarak açık görüşlü olmaya teşvik ederek doğruluğu artırabilirdi.

3. *Seçme:* Bütün tahminciler tahmin isabetlerine göre puanlandı ve tam bir yılın sonunda, en iyi %2'lik dilim süper tahminci ilan edilerek onlara ertesi yıl elit takımlarda birlikte çalışma fırsatı verildi.

Görüldü ki üç müdahale de kişilerin Brier skorlarını iyileştirmeleri açısından işe yaradı. Eğitim fark yarattı, takım olma daha büyük bir fark yarattı ve seçmenin ondan da büyük etkisi oldu.

Bu önemli bulgu yargıları bir araya toplamanın ve iyi karar vericiler seçmenin değerini doğruluyor. Ama hikâyenin tamamı bu değil. Her müdahalenin etkileri hakkında verilerle donanan ve Tetlock ve Mellers'la iş birliği yapan Ville Satopää her müdahalenin tahminleri tam olarak nasıl iyileştirdiğini derinlemesine incelemek için sofistike bir istatistik tekniği geliştirdi. Prensipte bazı tahmincilerin diğerlerinden daha iyi ya da daha kötü performans sergilemesinin üç ana nedeni olduğunu düşünüyordu:

1- Ortamda yapmaları gereken tahminle alakalı veri bulma ve analiz etme konusunda daha yetenekli olabilirler. Bu açıklama bilginin önemine dikkat çekiyor.

2- Bazı tahmincilerin bir tahminin gerçek değerinin belli bir tarafında hata yapmaya genel bir eğilimleri olabilir. Yüzlerce tahmin arasından, siz statükodan belirli değişiklilerin olması olasılığını sistematik olarak aşırı fazla ya da aşırı eksik tahmin ediyorsanız ya değişim ya da istikrar lehine bir ön yargı türünden muzdarip olduğunuz söylenebilir.

3- Bazı tahminciler gürültüye (ya da rastgele hatalara) daha az açık olabilirler. Tahminde, her yargıda olduğu gibi, gürültünün pek çok tetikleyicisi olabilir. Tahminciler belirli bir habere aşırı tepki verebilirler (bu örüntü gürültüsü dediğimiz şeyin bir örneğidir), durum gürültüsüne açık olabilirler ya da olasılık ölçeği kullanımlarında gürültülü olabilirler. Bütün bu gürültüler ve daha fazlası boyutları ve yönleri açısından öngörülemezdir.

Satopää, Tetlock, Mellers ve meslektaşları Marak Salikhov modellerine tahmin için BIN (ön yargı, bilgi ve gürültü) modeli adını verdiler. Üç ögenin her birinin üç müdahalenin her birindeki performans iyileşmesinden ne kadar sorumlu olduğunu ölçmek için kolları sıvadılar.

Cevap basitti: Üç müdahale de öncelikli olarak gürültüyü azaltarak işliyordu. Araştırmacıların ifadesiyle "Müdahaleler, isabeti hızlı bir şekilde güçlendirdikleri durumlarda, bunu yargıdaki rastgele hataları bastırarak yapıyorlardı. İlginç bir şekilde, eğitim müdahalesinin başlangıçtaki amacı ön yargıyı azaltmaktı."

Eğitim ön yargıları azaltmak için tasarlandığı için, süper olmayan bir tahminci ön yargı azaltmanın eğitimin en büyük etkisi olacağını öngörebilirdi. Ancak eğitim, gürültüyü azaltarak işe yarıyordu. Sürprizin açıklaması zor değildi. Tetlock'un eğitimi psikolojik ön yargılarla savaşmak için tasarlandı. Artık bildiğiniz gibi, psikolojik ön yargıların etkisi her zaman istatistiksel ön yargı değildir. Psikolojik ön yargılar, farklı yargılarda farklı bireyleri farklı şekillerde etkiledikleri zaman, gürültü üretirler. Tahmin edilen olaylar hayli çeşitli olduğu için, burada da durum bu. Aynı ön yargılar, konuya bağlı olarak, bir tahmincinin aşırı ya da gereğinden az tepki vermesine neden olabilir. Bir tahmincinin bir olayın gerçekleşip gerçekleşmeyeceğine genel eğilimi olarak tanımlanan *istatistiksel* ön yargı üretmelerini beklememeliyiz. Sonuç olarak tahmincileri psikolojik ön yargılarıyla savaşmak için eğitmek, gürültüyü azaltma yoluyla işe yarar.

Takım olmanın gürültüyü azaltma üzerindeki etkisi nispeten daha büyüktü ama aynı zamanda takımların bilgi çıkarma becerisini de önemli ölçüde iyileştiriyordu. Sonuç bir araya toplama mantığıyla tutarlı; birlikte çalışan birden fazla beyin bilgi bulmak konusunda tek başına çalışan bir beyinden daha iyi iş çıkarır. Alice ve Brian birlikte çalışıyorsa ve Alice Brian'ın kaçırdığı sinyalleri yakaladıysa ortak tahminleri daha iyi olacaktır. Süper tahminciler grup hâlinde çalıştıklarında grup kutuplaşmasının ve bilgi şelalelerinin tehlikelerinden kaçınmayı başarır gibi görünüyorlar. Bunlar yerine, verilerini ve içgörülerini bir havuzda topluyor ve aktif olarak açık görüşlü tarzlarıyla, bir araya toplanan bilgiden en fazla faydayı sağlıyorlar. Satopää ve meslektaşları bu avantajı

şöyle açıklıyor: "Takım olmak, eğitimin aksine, tahmincilerin bilgi toplamasına olanak sağlar."

En büyük toplam etkiyi seçme müdahalesi sağlıyordu. Sağladığı iyileşmenin bir kısmı bilginin daha iyi kullanılmasından geliyordu. Süper tahminciler ilgili bilgileri bulmak konusunda diğerlerinden daha iyiler. Muhtemelen daha akıllı, daha iyi motive ve bu tür tahminlerde bulunmak konusunda ortalama bir katılımcıdan daha tecrübeli oldukları için. Ancak seçmenin ana etkisi yine gürültüyü azaltmak. Süper tahminciler sıradan oyunculardan ve hatta eğitimli takımlardan daha az gürültülüler. Bu bulgu da Satopää ve diğer araştırmacılar için sürpriz oldu. "Süper tahminciler başarılarını haberlerin akıllıca okunmasından çok, ölçüm hatasını bastırmadaki üstün disiplinlerine borçlu olabilirler", ve bu diğerlerinin taklit edemeyeceği bir şeydir.

Seçme ve Bir Araya Toplamanın İşe Yaradığı Yerde

Süper tahmin projelerinin başarısı, iki karar hijyeni stratejisinin değerini vurguluyor: *seçme* (süper tahminciler, adı üstünde süperler) ve *bir araya toplama* (tahminciler takım hâlinde çalıştıklarında daha iyi performans sergilerler). İki strateji pek çok yargıda yaygın olarak uygulanabilirler. Mümkün olduğunda hem yaptıkları işte iyi olmak hem de birbirlerini tamamlamak için seçilmiş yargıç takımlar oluşturarak (tahminciler, yatırım profesyonelleri, işe alım memurları gibi) stratejileri birleştirmeyi hedeflemelisiniz.

Şu ana kadar kalabalıkların bilgeliği deneylerinde olduğu gibi çoklu bağımsız yargıların ortalamaları alınarak ulaşan iyileştirilmiş kesinliği değerlendirdik. Doğruluğu daha yüksek olan yargıçların tahminlerini bir araya toplamak isabetliliği daha da artıracaktır. Bununla birlikte, isabetlilikte bir başka kazanç hem bağımsız hem tamamlayıcı olan yargıları birleştirmekle sağlanabilir. Dört kişinin bir suça tanık olduğunu varsayalım. Elbette birbirlerini

etkilemediklerinden emin olmak önemlidir. Ek olarak, suçu dört farklı açıdan gördülerse, sağlayacakları bilginin kalitesi çok daha iyi olacaktır.

Birlikte karar almaları için bir profesyoneller takımı oluşturmak, adayların okuldaki ya da işteki gelecek performanslarını tahmin etmek için birçok testi bir araya toplamaya benzer. Bu görev için standart araç çoklu regresyondur (Bölüm 9'da değinmiştik). Değişkenleri peş peşe seçerek işler. Önce sonucu en iyi tahmin eden test seçilir. Bununla birlikte, dâhil edilecek bir sonraki test illaki ikinci en geçerli test olmak zorunda değildir. İlkine göre hem doğru hem de gereksiz olmayan tahminler sağlayarak birinci teste en çok kestirimci güç *ekleyecek* test olmalıdır. Örneğin, diyelim, gelecekteki performansla 0,50 ve 0,45 bağıntılı olan iki zihinsel kabiliyet testiniz ve performansla sadece 0,30 bağıntılı olan ama kabiliyetle bağıntısı olmayan bir kişilik testiniz var. Optimal çözüm önce daha doğru olan kabiliyet testini, daha sonra da yeni bilgi getiren kişilik testini seçmektir.

Benzer bir şekilde, bir yargıçlar takımı oluşturuyorsanız, elbette önce en iyi yargıcı seçmelisiniz. Ancak bir sonraki tercihiniz birinci yargıca çok benzeyen daha geçerli bir yargıçtan çok, orta düzeyde geçerli ama masaya yeni beceriler getiren bir aday olmalı. Bu şekilde seçilen bir takım daha üstün olacaktır çünkü bir araya toplanan yargıların geçerliliği, yargılar birbirleriyle ilişkisiz olduğunda, gereksiz oldukları duruma göre daha hızlı artar. Örüntü gürültüsü böyle bir takımda nispeten yüksek olacaktır çünkü her yargının bireysel yargısı farklılık gösterecektir. Paradoksal bir şekilde, o gürültülü grubun ortalaması, görüş birliğinin sağlandığı bir grubun ortalamasından daha isabetli olacaktır.

Önemi bir uyarıda bulunmalıyız. Çeşitlilikten bağımsız olarak, bir araya toplama gürültüyü ancak yargılar gerçek anlamda bağımsız olduğunda azaltabilir. Gruplarda gürültüden bahsederken vurguladığımız gibi, grup müzakeresi sıklıkla gürültüdeki hatayı

azalttığından çok, ön yargıya hata ekler. Çeşitliliğin gücünden faydalanmak isteyen örgütler takım üyelerinin bağımsız olarak yargıya varmaları durumunda doğacak anlaşmazlıklara açık olmalılar. Hem bağımsız hem çeşitli olan yargıları meydana çıkarmak ve bir araya toplamak, sıklıkla en kolay, en ucuz ve en yaygın olarak uygulanabilen karar hijyeni stratejisi olacaktır.

Seçme ve Bir Araya Toplamadan Bahsederken

"Dört bağımsız yargının ortalamasını alalım; gürültüyü yarı yarıya düşürmesi garantilidir."

"Süper tahminciler gibi, sürekli beta hâlinde olmak için çabalamalıyız."

"Bu durumu tartışmadan önce, ilgili baz oranı nedir?"

"İyi bir takımımız var ama daha fazla görüş çeşitliliğini nasıl sağlayabiliriz?"

BÖLÜM 22

Tıpta Kılavuzlar

Yıllar önce iyi bir dostumuza (adına Paul diyelim) aile hekimi tarafından yüksek tansiyon teşhisi kondu (doktora da Dr. Jones diyelim). Doktor Paul'e ilaç kullanmayı denemesini önerdi. Bir idrar söktürücü reçete etti ama bir etkisi olmadı. Paul'ün tansiyonu yüksek kalmaya devam etti. Birkaç hafta sonra, Dr. Jones bu duruma ikinci bir ilaçla, bir kalsiyum kanal blokeriyle tepki verdi. Bu ilacın etkisi de hayli mütevazıydı.

Bu sonuçlar Dr. Jones'u şaşırttı. Haftalık muayenehane ziyaretleriyle geçen üç ayın sonunda, Paul'ün yüksek tansiyon değerleri hafif bir düşüş gösterse de hâlâ yüksekti. Sonraki adımların ne olacağı net değildi. Paul kaygılı, Dr. Jones sıkkındı ve bunun en önemli nedenlerinden biri Paul'ün sağlıklı ve nispeten genç bir adam olmasıydı. Dr. Jones üçüncü bir ilaç denemeyi düşünüyordu.

O noktada Paul tesadüfen yeni bir şehre taşındı ve yeni bir aile hekimine muayene oldu (adına Dr. Smith diyeceğiz). Paul Dr. Smith'e yüksek tansiyonla bitmeyen mücadelesini anlattı. Dr. Smith derhâl, "Bir tansiyon aleti edinin, değerlere bakalım. Ben yüksek tansiyon hastası olduğunuzu hiç sanmıyorum. Muhtemelen sizde

beyaz önlük sendromu var; tansiyonunuz doktor muayenehanelerinde yükseliyor!" dedi.

Paul söyleneni yaptı ve tahmin edildiği gibi, evdeyken tansiyonu normaldi. O zamandan beri hâlâ normal (ve Dr. Smith'in beyaz önlük sendromundan bahsetmesinden bir ay sonra, muayenehanede de normale indi).

Doktorların temel görevi teşhis koymak, bir hastanın hastalığı olup olmadığına karar vermek ve eğer hastaysa hastalığını teşhis etmektir. Teşhis sıklıkla bir yargıyı gerektirir. Pek çok şartta, teşhis rutin ve hayli mekaniktir ve gürültüyü azaltmak için birtakım kural ve prosedürler uygulanır. Bir doktorun bir kişinin omzunun çıkıp çıkmadığını ya da ayak parmağını kırıp kırmadığını saptaması genellikle kolaydır. Benzer bir şey daha teknik sorunlar için de söylenebilir. Tendon dejenerasyonunun miktarını belirlemek ise biraz gürültü üretir. Patologlar meme lezyonlarının kalın iğne biyopsilerini değerlendirirken, değerlendirmeleri nispeten az gürültülü ve açıktır.

Bazı teşhisler hiç yargı içermez. Sağlık bakım sıklıkla yargı ögesini ortadan kaldırarak, yargıdan hesaba kayarak ilerler. Ağrılı boğaz vakasında, doktor işe hastanın gırtlağından alınan sürüntü örneğine hızlı bir antijen testi uygulayarak başlar. Kısa bir süre içinde, test enfeksiyonu saptayacaktır. (Hızlı antijen sonucu olmadığında ve hatta bir dereceye kadar test sonucu olduğunda bile, boğaz enfeksiyonu teşhisinde bir miktar gürültü olabilir.) Açlık kan şekeriniz 126 mg/dL ya da daha yüksekse ya da HbA1c testi (önceki üç ay boyunca kan şekeri ölçümünün ortalaması) en az 6,5 ise diyabetli kabul edilirsiniz. COVID-19 pandemisinin ilk aşamalarında, bazı doktorlar semptomları değerlendirdikten sonra varılan yargıların bir sonucu olarak teşhis koyuyordu. Pandemi ilerledikçe testler yaygınlaştı ve yargıyı gereksiz kıldı.

Pek çok insan doktorların yargı kullanırken gürültülü olabildiğini ve hata yapabildiklerini bilir: Hastalar için standart bir

pratik ikinci bir görüş alınmasını tavsiye etmektir. Hatta bazı hastanelerde ikinci görüş zorunludur. İkinci görüş ilkiyle uyuşmadığında, gürültü var demektir. Gerçi elbette hangi doktorun haklı olduğu net olmayabilir. Bazı hastalar (Paul gibi) ikinci görüşün ilkinden ne kadar farklı olduğuna şaşırırlar. Ama sürpriz olan tıp mesleğinde gürültünün varlığı değil, büyüklüğüdür.

Bu bölümdeki amaçlarımız bu iddiayı masaya yatırmak ve tıp mesleğinde gürültüyü azaltmaya yönelik yaklaşımların bazılarını anlatmak olacak. Karar hijyeni stratejisine, teşhis kılavuzlarının geliştirilmesine odaklanacağız. Tıpta gürültü ve doktorların, hemşirelerin ve hastanelerin çare bulmak için attıkları adımlar konusunda koca bir kitabın yazılabileceğinin fazlasıyla farkındayız. Dikkate değer bir şekilde, tıpta gürültü burada odağımız olan teşhis yargılarındaki gürültüyle sınırlı değildir. Tedaviler de gürültülü olabilir ve çok kapsamlı bir literatür bu konuyu da ele almaktadır. Bir hastanın kalp sorunu varsa, konu ister doğru ilaç, ister doğru ameliyat ya da ameliyatın gerekli olup olmaması olsun, doktorların en iyi tedavi konusundaki yargıları şaşırtıcı farklılıklar gösterebilir. Dartmouth Atlas Projesi yirmi yılı aşkın süredir kendini "ABD'de tıbbi kaynakların nasıl dağıldığı ve kullanıldığı konusundaki çarpıcı farklılıkları" belgelemeye adadı. Çok çeşitli ülkelerde benzer sonuçlar gözleniyor. Ancak, amaçlarımız doğrultusunda, teşhis yargılarındaki gürültünün kısa bir keşfi yeterli olacaktır.

Ufukta Bir Tur

Tıpta gürültü üzerine çok geniş bir literatür var. Literatürün büyük kısmı gürültünün varlığını test eden gözlem ve deneye dayalı olmakla birlikte, büyük bir kısmı da kural koyucudur. Sağlık bakımla ilgili olan uzmanlar, gürültü azaltma stratejileri aramaya

devam ediyorlar ve pek çok şekle girebilen bu stratejilerin pek çok alanda değerlendirilmeye layık bir fikir madenî olduğu söylenebilir.

Gürültü olduğunda bir hekim açıkça haklıyken diğer açıkça hatalı (ve ön yargının bir türünden muzdarip) olabilir. Beklenece̊i gibi, yetenekler çok önemlidir. Örneğin, radyologların zatürre teşhisleri üstünde yapılan bir incelemede, önemli ölçüde gürültü bulundu. Bu gürültünün büyük kısmı beceri farklılıklarından kaynaklanıyordu. Daha detaylı ifade etmek gerekirse, "beceri farklılıkları teşhis kararlarındaki farklılıkların %44'ünü açıklayabilir" ve "beceriyi iyileştiren politikalar tek tip karar kılavuzlarından daha iyi performans sergileyebilirler." Başka alanlarda olduğu gibi burada da, eğitim ve seçme hatanın azaltılması ve hem gürültünün hem ön yargının ortadan kaldırılması açısından hayati önem taşır.

Radyoloji ve patoloji gibi bazı uzmanlık alanlarında, doktorlar gürültünün varlığının farkındadırlar. Örneğin radyologlar teşhis farklılıklarını "Aşil topuğu" olarak adlandırırlar. Radyoloji ve patoloji alanlarındaki gürültünün bu alanlarda diğerlerinden daha fazla gürültü olduğu için mi, yoksa gürültü burada daha kolay belgelenebildiği için mi dikkat çektiği çok açık değil. Belgeleme kolaylığının önemli olduğundan şüpheliyiz. Temiz, basit gürültü (ve bazen hata) testlerini radyolojide yapmak daha kolaydır. Örneğin daha önceki değerlendirmeleri yeniden gözden geçirmek için taramalara ve slaytlara dönmek mümkündür.

Tıpta kişiler arası gürültü ya da *değerlendiriciler arası güvenilirlik* genellikle *kappa istatistiğiyle* ölçülür. Kappa ne kadar yüksekse gürültü o kadar azdır. 1'lik bir kappa değeri kusursuz anlaşmayı yansıtır, 0'lık bir kappa değeri ise ancak bir olası teşhisler listesine dart artan maymunlardan bekleyebileceğiniz kadar anlaşma demektir. Bazı tıbbi teşhis alanlarında, bu katsayıyla ölçülen güvenilirliğin "yetersiz" ya da "az" olduğu görülmüştür ve bu, gürültünün çok yüksek olduğu anlamına gelir. Güvenirlik sıklıkla "makul" olarak değerlendirilir ki bu çıkarım elbette daha iyi

olmakla birlikte önemli miktarda gürültüye işaret eder. Hangi ilaç etkileşimlerinin klinik olarak anlamlı olduğu konusunda yapılan bir çalışmada, rastgele seçilmiş yüz ilaç etkileşimini değerlendiren pratisyen hekimler arasında "yetersiz anlaşma" görülmüştür. Konunun dışındaki kişilere ve pek çok doktora böbrek hastalıklarının farklı evrelerinin teşhisi nispeten dolambaçsız ve açık görünebilir. Fakat nefrologlar arasında, böbrek hastalığına sahip hastaların değerlendirilmesinde kullanılan standart testlerin anlamı konusundaki yargılarında sadece "az ve orta anlaşma" gözlenmektedir.

Bir meme lezyonunun kanserli olup olmadığı konusunda yapılan bir çalışmada, patologlar arasında sadece "makul" düzeyde anlaşma saptandı. Memede proliferatif lezyon teşhislerinde, görüş birliği yine "makul" idi. Spinal stenozis derecesini saptamak için MRI sonuçlarını inceleyen doktorlar arasında da görüş birliği yine "makul" düzeydeydi. Bu bulgular karşısında durup düşünmekte fayda var. Bazı alanlarda tıpta gürültü seviyesinin çok düşük olduğundan bahsetmiştik. Ancak hayli teknik olan bazı alanlarda doktorlar gürültüsüz olmaya çok uzaklar. Hastaya kanser gibi ciddi bir hastalık teşhisi konup konmayacağı bile, hangi doktora görüneceğinin belirlediği bir tür piyangoya bağlı olabilir.

Gürültü miktarının özellikle dikkate değer göründüğü alanlardan çekilmiş birkaç bulguyu daha değerlendirin. Bu bulguları gelişmeye ve ilerlemeye devam eden (bazı örneklerde hızla) tıp pratiğinin mevcut durumu hakkında otoriter beyanlarda bulunmak için değil, gürültünün hem nispeten yakın geçmişte hem de bugün ne kadar yaygın olduğu konusunda genel bir hissi aktarmak için paylaşıyoruz.

1- Kalp hastalıkları ABD'de hem erkeklerde hem kadınlarda başlıca ölüm nedeni. Kalp hastalıklarının test edilmesinde kullanılan başlıca yöntem olan koroner anjiyogramlar hem akut hem akut olmayan ortamlarda kalp arterlerindeki tıkanıklık derecesini değerlendirir. Akut olmayan ortamlarda, bir hasta tekrarlı göğüs

ağrısı şikâyetiyle geldiğinde, tedavi –stent takılması gibi– genellikle arterlerden birinin ya da daha fazlasının %70'ten fazla tıkalı olması hâlinde uygulanır. Ancak anjiogramların yorumlanmasında gereksiz prosedürlere yol açma potansiyeli olan farklılıklar belgelenmiştir. Eski bir çalışma vakaların %31'inde anjiyogramı değerlendiren hekimlerin bir ana damarın %70'in üstünde tıkalı olup olmaması konusunda görüş ayrılığı yaşadıklarını göstermiştir. Kardiyologlar arasında anjiyogramların yorumlanmasında farklılık yaşandığının bilinmesine ve süregelen çaba ve düzeltici adımlara rağmen, sorun henüz çözülebilmiş değildir.

2- Endometriozis, normalde rahmin içini kaplayan endometriyum dokusunun rahmin dışında büyüdüğü bir rahatsızlıktır. Bu bozukluk ağrıya ve kısırlık sorunlarına yol açabilir. Genellikle küçük bir kameranın cerrahi yollardan vücuda sokulduğu laparoskopi yöntemiyle teşhis edilir. İkisinde farklı derecelerde endometriozis olan, birinde ise hiç olmayan üç hastanın laparoskopi videoları 108 jinekolojik cerraha gösterildi. Cerrahlardan endometriotik lezyonların sayısını ve konumlarını değerlendirmeleri istendi. Cerrahlar hem sayı hem konum açısından çok zayıf korelasyonlarla, çarpıcı görüş farklılıkları sergilediler.

3- Tüberküloz (TB) dünya çapında en yaygın ve ölümcül hastalıklardan biridir. Sadece 2016 yılında, 10 milyondan fazla insan tüberküloza yakalandı ve neredeyse 2 milyon kişi hayatını kaybetti. Tüberkülozu saptamak için yaygın olarak kullanılan yöntemlerden biri, TB bakterisinin ciğerlerde açtığı boş alanların incelenmesine olanak sağlayan göğüs röntgenidir. TB teşhisindeki değişkenlik neredeyse yetmiş beş yıldır kayıt altına alınıyor. On yıllar içinde yaşanan iyileşmelere rağmen, çalışmalar TB teşhislerinde, değerlendiriciler arasında "orta" ya da "makul" görüş birliğiyle sınırlı kalan önemli farklılıklar saptamaya devam ediyor. Ayrıca TB teşhislerinde farklı ülkelerin radyologları arasında da değişkenlikler mevcut.

4- Patologlar melanom için (en tehlikeli cilt kanseri türü) deri lezyonu incelemelerinde sadece "orta" düzeyde görüş birliğine ulaştılar. Her vakayı gözden geçiren sekiz patolog aynı görüşteydi ve örneklerin %62'sinde sadece bir anlaşmazlık yaşadılar. Bir onkoloji merkezinde yapılan bir başka çalışmada, melanomların teşhis isabetliliği sadece %64'tü; yani doktorlar üç lezyondan birine yanlış teşhis koyuyordu. Üçüncü bir çalışma ise New York Üniversitesi'nde dermatologların deri biyopsilerinde melanom teşhislerinin %36'sında yanıldıklarını ortaya koydu. Çalışmanın yazarları "melanomun klinik olarak doğru bir şekilde teşhis edilememesinin potansiyel olarak ölümcül olan bu hastalığa sahip olan hastaların hayatta kalmalarında çok ciddi etkisi" olduğu sonucuna vardılar.

5- Radyologların meme kanserinde mamografi incelemelerine dayanan yargılarında da büyük farklılıklar var. Büyük bir çalışma, farklı radyologlar arasında hatalı olumsuz saptamaların oranının %0 ile (radyolog her defasında haklıydı) %50'nin üstü (radyolog incelemelerin yarısında mamografi görüntüsünü hatalı bir şekilde normal olarak yorumladı) arasında değiştiğini ortaya koydu. Benzer bir şekilde hatalı olumlu değerlendirmeler de %1 ile %64 arasında değişiyordu (yani incelemelerin neredeyse üçte ikisinde radyolog ortada kanser olmamasına rağmen, mamografi sonucunun kanseri işaret ettiğini düşündü). Farklı radyologlardan gelen hatalı olumsuzlar ve hatalı olumlular ortada gürültü olduğunu doğruluyordu.

Kişiler arası gürültü vakaları mevcut araştırmalarda daha baskın olmakla birlikte durum gürültüsü bulguları da yok değil. Radyologlar bazen aynı imgeyi tekrar değerlendirirken farklı bir görüş sunuyor ve dolayısıyla kendi kendileriyle görüş ayrılığı yaşıyorlar (başkalarıyla yaşadıklarından daha nadir olsa da). Anjiyogramlardaki tıkanıklık derecesini değerlendirirken, yirmi iki hekim %62 ve %92 arası oranlarda kendileriyle çeliştiler. Belirsiz

kriterleri ve karmaşık yargıları içeren alanlarda, değerlendiriciler arası güvenilirlik hayli yetersiz olabiliyor.

Bu çalışmalar bu ortam gürültüsüne net bir açıklama sağlamıyor. Ancak teşhisleri içermeyen bir başka çalışma, tıpta basit bir durum gürültüsü kaynağını hem hastalar hem doktorlar için üstünde durmaya değer bir bulguyu saptadı. Kısaca doktorlar sabahın erken saatlerinde kanser taraması talep etmeye öğleden sonradan daha yakındılar. Büyük bir örnek grubunda meme ve kolon tarama testi talepleri sabah saat 08.00'de %63,7 ile en yüksek seviyedeydi. Saat ilerledikçe azalarak 11.00'de %48,7'ye düşüyordu. Öğle saatlerinde yeniden %56,2'ye çıktıktan sonra saat 17.00'de %47,8'e düşüyordu. Bir başka deyişle, günün daha ileri saatlerinde randevusu olan hastaların kılavuz tavsiyeli kanser taraması alma ihtimali daha düşüktü.

Bu tür bulguları nasıl açıklayabiliriz. Olası bir cevap doktorların her zamanki yirmi dakikalık süreden daha fazlasını gerektiren karmaşık tıbbi problemli hastaları gördükten sonra, programda kaçınılmaz olarak geride kalmaları. Daha önce durum gürültüsünün tetikleyicisi olarak stres ve yorgunluğun rolünden bahsetmiştik (bakınız Bölüm 7) ve bu ögeler burada da işbaşında görünüyor. Bazı doktorlar programa yetişmek için önleyici sağlık tedbirleriyle ilgili konuşmaları atlayabiliyorlar. Yorgunluğun hekimler arasındaki bir diğer rolü de hastane vardiyalarının sonunda el yıkama oranının olması gerekenin altında kalması. (El yıkamanın da gürültülü olduğu anlaşılıyor.)

Daha Az Gürültülü Doktorlar: Kılavuzların Değeri

Farklı tıbbi sorunlar bağlamında, gürültünün varlığı ve büyüklüğü konusunda kapsamlı bir açıklama sağlamak, sadece tıbba değil insan bilgisine de çok büyük bir katkı olurdu. Bu tür bir açıklamanın varlığından habersiziz; gelecekte üretileceğini umuyoruz. Ama şimdi bile, var olan bulgular birtakım ipuçları veriyor.

Bir uçta, bazı sorun ve hastalıkların teşhisi özünde hayli mekanik ve yargıya yer bırakmıyor. Diğer vakalarda ise, teşhis mekanik değil ama düz; tıp eğitimi olan herkesin aynı sonuca varması kuvvetle muhtemel. Yine başka vakalarda –mesela akciğer kanseri uzmanları arasında– gürültünün var olduğundan ama en alt seviyede tutulduğundan emin olmak için belli bir derecede uzmanlık gerekir. Diğer uçta ise, bazı vakalar yargıya çok büyük yer bırakır ve teşhis için alakalı kriterler o kadar açık uçludur ki gürültü azımsanmayacak kadar büyük ve azaltılması güç olur. Daha sonra göreceğimiz gibi, psikiyatrinin büyük kısmında bu durum geçerlidir.

Tıpta gürültüyü azaltmada ne işe yarayabilir? Daha önce bahsettiğimiz gibi, eğitim beceriyi artırır ve becerinin faydası mutlaka olur. Çoklu uzman yargılarının bir araya toplanması da öyle (ikinci görüşler vs.). Algoritmalar özellikle umut vadeden bir yoldur ve doktorlar gürültüyü azaltmak için derin öğrenme algoritmaları ve yapay zekâ kullanmaya başladılar. Örneğin, bu tür algoritmalar meme kanseri olan kadınlarda lenf modu metastazlarını saptamak için kullanılıyor. Bunların en iyilerinin en iyi patologlardan daha üstün olduğu görüldü ve elbette algoritmalar gürültülü değildir. Ayrıca diyabetle ilişkilendirilen göz sorunlarının saptanmasında derin öğrenme algoritmaları da hatırı sayılır bir başarıyla kullanılıyor. Ve YZ artık mamografilerden kanser saptama konusunda en az radyologlar kadar iyi iş çıkarıyor; YZ'nin gelecekteki ilerlemeleri muhtemelen onu da üstün kılacaktır.

Tıp mesleğinin gelecekte algoritmalara gittikçe daha fazla bel bağlaması çok mümkün görünüyor; algoritmalar hem ön yargıyı hem gürültüyü azaltmayı ve bu süreçte hayat kurtarmayı ve tasarruf sağlamayı vadediyorlar. Ancak bizim buradaki vurgumuz insan yargı kılavuzları olacak çünkü tıp alanı kılavuzların bazı uygulamalarda nasıl iyi ve hatta kusursuz sonuçlar, diğerlerinde ise nasıl daha karışık sonuçlar ürettiğini çok iyi gösteren bir örnek.

Belki de teşhis kılavuzlarına en iyi örnek, 1952 yılında, obstetrik anestezi uzmanı Virginia Apgar tarafından geliştirilen Apgar puanlamasıdır. Bir yenidoğanın sıkıntıda olup olmadığının değerlendirilmesi doktorlar ve ebeler için bir klinik yargı meselesiydi. Apgar puanlaması onlara standart bir kılavuz sağladı. Değerlendirici, bebeğin rengini, nabzını, reflekslerini, kas sıkılığını, solunum çabasını ölçüyor. Bunlar bazen İngilizce karşılıklarının baş harfleri Apgar'ın adını oluşturacak şekilde özetlenirler: *Görüntü, (Appereance-* ten rengi), *nabız (Pulse,* kalp atışı), *yüz buruşturma (Grimace,* refleksler), *aktivite (Activity,* kas sıkılığı) ve *solunum (Respiration,* solunum hızı ve harcanan çaba). Apgar testinde bu beş ölçümün her birine 0, 1 ya da 2 puan verilir. En yüksek toplam puan 10 olabilir ve nadirdir. 7 ve üstü iyi sağlık göstergesi olarak kabul edilir. (Tablo 3)

Tablo 3: *Apgar Puanlama Kılavuzu*

Kategori	*Verilen puanlar*
Görüntü	0: Bütün vücut mor ya da solgun
(ten rengi)	1: Vücutta renk iyi ama eller ya da ayaklar mor
	2: Tamamen pembe ya da normal renkte
Nabız	0: Nabız yok
(kalp atışı)	1: < dakikada 100 atış
	2: > dakikada 100 atış
Yüz buruşturma	0: Hava yollarının uyarılmasına hiç tepki yok
(refleksler)	1: Uyarılmaya karşı yüzü buruşturma
	2: Uyarılmaya karşı yüzü buruşturma ve öksürme ya da hapşırma
Aktivite	0: Cansız
(kas sıkılığı)	1: Kol ve bacakların bir miktar esnetilmesi (kıvrılması)
	2: Aktif hareket
Solunum	0: Nefes almıyor
(solunum hızı ve	1: Cılız ağlama (sızlanma, homurtu)
çaba)	2: İyi ve güçlü ağlama

Kalp atışının puanlama sisteminin tek sayısal ögesi olduğuna ve diğer maddelerin hepsinin bir miktar yargı içerdiğine dikkat edin. Ancak yargılar, her biri değerlendirmesi çok açık olan bireysel ögelere ayrıldığı için, mütevazı bir eğitim seviyesindeki pratisyenlerin bile büyük miktarda görüş ayrılığı yaşaması çok olası değil; bu nedenle Apgar puanlamasının çok az gürültü ürettiğini söyleyebiliriz.

Apgar puanlaması kılavuzların nasıl işlediğine ve gürültüyü neden azalttıklarına iyi bir örnektir. Kuralların ve algoritmaların aksine kılavuzlar yargı ihtiyacını ortadan kaldırmazlar; karar düz bir hesaplamadan ibaret değildir. Ögelerin her birinde ve dolayısıyla nihai kararda anlaşmazlık mümkündür. Bununla birlikte, kılavuz gürültüyü azaltmakta başarılı olur çünkü karmaşık bir kararı, önceden belirlenmiş boyutlarda bir dizi daha basit alt yargıya çözümler.

Bu yaklaşımın getirileri, soruna Bölüm 9'da bahsi geçen basit öngörü modelleri açısından baktığımız zaman netlik kazanır. Yeni doğan bir bebeğin sağlığı hakkında yargıda bulunan bir doktor çok sayıda kestirimci ipucuna göre hareket etmektedir. Durum gürültüsü söz konusu olabilir; hekim bir gün ya da bir ruh hâlinde nispeten daha az önemli kestirimcilere dikkat edip daha önemli olanları yok sayarken, başka bir gün ve ruh hâlinde tam tersi olabilir. Apgar puanlaması sağlık profesyonelini önemli oldukları deney ve gözleme dayalı olarak bilinen beş tanesine odaklar. Ardından, puan her işaretin nasıl değerlendirilmesi gerektiğinin açık bir tarifini verir ve bu, işaret düzeyindeki her yargıyı büyük ölçüde basitleştirerek gürültüsünü de azaltır. Son olarak Apgar puanlaması istenen toplam yargıyı üretmek için kestiricileri mekanik olarak nasıl tartmak gerektiğini de belirler, oysa insan sağlık profesyonelleri bu işaretlere atfettikleri ağırlıklarda çeşitlilik sergileyebilirler. İlgili kestiricilere odaklanmak, kestirimci modelin basitleştirilmesi ve mekanik bir araya toplama; bunların hepsi gürültüyü azaltır.

Tıbbın pek çok farklı alanında benzer yaklaşımlar kullanılmıştır. Bunlardan biri, boğaz enfeksiyonunun teşhisine rehberlik eden Centor puanlamasıdır. Hastaya aşağıdaki semptomların ya da işaretlerin her biri için bir puan verilir (Bu terimlerin baş harfleri de tıpkı Apgar puanlamasında olduğu gibi, bu kılavuzu meslektaşlarıyla birlikte ilk kez özetleyen Robert Centor'un soyadını oluşturur): *öksürük (Cough)* yokluğu, *Eksüda* (boğazın arka tarafında beyaz lekeler) varlığı, boyunda hassas ya da şişkin lenf *Nodları* ve 38 derecenin üstüne çıkan ateş *(Temperature).* Hastaya verilen puana bağlı olarak, streptokok farenjit teşhisi için sürüntü örneği önerilebilir. Bu ölçek kullanıldığında, değerlendirme ve puanlama hayli açıktır ve bu durum gereksiz test yapılan ve streptokok farenjit tedavisi gören insanların sayısını azaltmıştır.

Benzer şekilde, mamografilerin yorumlanmasında gürültüyü azaltan BI-RADS (Meme Görüntüleme Raporlama ve Veri Sistemi) ile meme kanseri teşhisi için de kılavuz belirlenmiştir. Bir çalışma, BI-RADS'ın mamografilerin değerlendirilmesinde değerlendiriciler arası görüş birliğini artırdığını ortaya koymuştur. Bu da bize kılavuzların, farklılıkların çok fazla olduğu bir alanda gürültünün etkili olabileceğini göstermektedir. Patolojide de kılavuzları aynı amaçla kullanmak için birçok başarılı çaba sergilenmektedir.

Can Sıkıcı Psikiyatri Örneği

Gürültü açısından, psikiyatri en uç örnektir. Psikiyatrlar aynı hastaya aynı teşhis kriterlerini kullanarak teşhis koyarken bile birbirleriyle sık sık görüş ayrılığı yaşarlar. Bu nedenle, gürültüyü azaltmak en azından 1940'lardan beri psikiyatri toplumu için önemli bir öncelik olmuştur. Ve göreceğimiz gibi, sürekli yenilenmelerine rağmen, kılavuzlar gürültüyü azaltmak konusunda sadece çok küçük bir yardım sağlayabilmiştir.

1964 yılında 91 hasta ve tecrübeli on psikiyatrın katıldığı bir çalışma, iki görüş arasında anlaşma sağlanması olasılığının sadece %57 olduğunu ortaya koydu. İki ayrı psikiyatr tarafından bağımsız olarak teşhis konulan 426 devlet hastanesi hastasını içeren bir başka çalışmada ise, var olan hastalığın türü konusunda, vakaların sadece %50'sinde görüş birliği gözlendi. Yine 153 ayakta tedavi gören hastanın konu olduğu bir diğer çalışma, %54 oranında görüş birliği saptadı. Bu çalışmalarda, gürültünün kaynağı belirtilmemişti. Ancak ilginç bir şekilde, bazı psikiyatrların hastaları belli teşhis kategorilerine sokmaya meyilli olduğu görülmüştü. Örneğin bazı psikiyatrlar hastalarına depresyon, diğerleri de anksiyete teşhisi koymaya özellikle meyilliydi.

Birazdan göreceğimiz gibi, psikiyatride gürültü seviyeleri yüksek seyrini koruyor. Neden? Uzmanlar tek bir net cevap bulamıyorlar (Bu da gürültünün izahının da gürültülü olduğu anlamına geliyor). Hiç şüphesiz, teşhis kategorilerinin çok fazla olması önemli bir faktör. Ancak araştırmacılar bu soruyu cevaplamak için bir ön çaba olarak bir hastayla önce bir psikiyatrın görüşme yapmasını istediler. Ardından, kısa bir dinleme molasını takiben, ikinci bir psikiyatr bir görüşme daha gerçekleştirdi. Sonrasında iki psikiyatr bir araya geldiler ve görüş ayrılığı yaşıyorlarsa, nedenini tartıştılar.

Sık rastlanan bir neden "hekimin değişkenliği" oldu: farklı düşünce ekolleri, farklı eğitimler, farklı klinik tecrübeler, farklı mülakat tarzları. "Gelişimsel eğitimli bir hekim halüsinasyon tecrübesini geçmiş bir istismarın travma sonrası tecrübesi" olarak açıklayabilirken, "biyomedikal oryantasyonu olan" bir başkası "aynı halüsinasyonları şizofrenik bir sürecin parçası" olarak yorumlayabilir. Bu tür farklılıklar örüntü gürültüsünün örnekleridir.

Doktor farklılıklarının ötesinde, gürültünün ana nedeni "terimlendirme yetersizliği" idi. Benzeri gözlemler ve psikiyatrik terimlendirme konusunda yaygın profesyonel tatminsizlik *Zihinsel Bozuklukların Tanısal ve Sayımsal El Kitabı* (DSM-III)'ün 1980'de

elden geçirilmesini (üçüncü baskı) motive etti. El kitabına ilk kez zihinsel bozuklukları teşhis etmek için açık ve detaylı kriterler dâhil edildi. Bu, bir teşhis kılavuzunun hazırlanması yönünde ilk adım oldu.

DSM-III teşhislerin gürültülü olup olmadığı yönündeki araştırmalarda çarpıcı bir artışa yol açtı. Ayrıca gürültüyü azaltmak konusunda da yardımcı oldu. Ancak el kitabı tam bir başarı olmaya çok uzaktı. DSM-IV'ün 2000 yılındaki hatırı sayılır revizyonundan sonra bile (ilk hâli 1994'te basılmıştı), araştırmalar gürültü seviyesinin hâlâ yüksek olduğunu gösteriyordu. Bir yandan Ahmed Aboraya ve meslektaşları "psikiyatrik bozukluklar için teşhis kriterlerinin kullanılmasının teşhislerin güvenilirliğini artırdığı" saptamasında bulundu. Öte yandan, "tek bir hastanın kayıtlarının bile, aynı hasta için çok sayıda teşhisi açığa çıkarması" riski ciddiyetini hâlâ koruyor.

El kitabının bir diğer versiyonu, DSM-5 2013 yılında piyasaya sürüldü. Amerikan Psikiyatri Derneği DSM-5'in gürültüyü azaltacağını ummuştu çünkü yeni baskı, daha objektif ve açıkça ölçeklenmiş kriterlere dayanıyordu. Ancak psikiyatrlar hâlâ önemli ölçüde gürültü sergiliyorlar. Örneğin, Samuel Lieblich ve çalışma arkadaşları "psikiyatrların kimde majör depresif bozukluk olduğu ve olmadığı konusunda hemfikir olmakta sıkıntı çektiklerini" belirtiyorlar. DSM-5 için saha denemeleri "asgari düzeyde görüş birliği" saptadı. Bu, "yüksek eğitimli uzman psikiyatrların çalışmanın şartları altında bir hastanın depresyonda olduğu konusunda, vakaların sadece %4 ila %15'i arasında görüş birliği sağlayabildikleri" anlamına geliyor. Birtakım saha denemelerine göre, DSM-5 durumu daha da kötüleştirerek "başlıca alanların hepsinde, karışık anksiyete ve depresif bozukluk gibi klinik uygulamada işe yaramayacak kadar güvenilmez teşhislerle gürültünün artmasına neden oldu."

Kılavuzların sınırlı başarısının başlıca nedeni, psikiyatride, "bazı bozuklukların teşhis kriterlerinin hâlâ müphem ve kulla-

nımı zor olması" gibi görünüyor. Bazı kılavuzlar gürültüyü yargıyı anlaşmazlığın azaldığı kriterlere bölerek azaltır ama bu tür kriterlerinin nispeten açık uçlu olması hâlinde, gürültü olasılığı sürer. Bu noktayı aklımızın bir köşesinde tutarak, öne çıkan tekliflerin daha standardize teşhis kılavuzlarına çağrıda bulunduğunu belirtelim. Bu teklifler (1) teşhis kriterlerini netleştirerek muğlak standartlardan uzaklaşmayı, (2) "hekimlerin semptomların varlığı ya da yokluğu konusunda görüş birliğine vardıklarında, teşhis konusunda hemfikir olmaya daha yatkın olmaları" teorisinden yola çıkarak, semptomların ve ciddiyet seviyelerinin "referans tanımlarını" üretmeyi ve (3) açık sohbetlere ek olarak hastalarla yapılandırılmış mülakatlar kullanmayı içerir. Tavsiye edilen bir mülakat rehberi örneğin, anksiyete, depresyon ve yeme bozukluklarının daha güvenilir teşhisine olanak sağlayan yirmi dört tarama sorusu içermektedir.

Bu adımlar umut vadeder gibi görünüyor ama gürültüyü azaltmak konusunda ne derece başarılı olacakları açık bir sorudur. Bir gözlemcinin sözcükleriyle, "hastanın öznel semptomlarına, hekimin semptomları yorumlamasına bel bağlanması ve nesnel ölçümün (kan testi gibi) yokluğu, psikiyatrik bozukluklar için güvensiz teşhislerin tohumlarını atmaktadır." Bu anlamda, psikiyatrinin gürültüyü azaltma girişimlerine özellikle dirençli olduğu söylenebilir.

Bu soruda, güvenli bir tahminde bulunmak için henüz çok erken. Ancak çok açık olan bir şey var. Genel olarak tıpta kılavuzlar hem ön yargıyı hem gürültüyü azaltmak konusunda çok başarılı oldular. Doktorlara, hemşirelere ve hastalara yardımcı oldular ve bu süreçte kamu sağlığını büyük ölçüde iyileştirdiler. Tıp mesleğinin daha fazla kılavuza ihtiyacı var.

Tıpta Kılavuzlardan Bahsederken

"Doktorlar arasında, gürültü düzeyi tahmin ettiğimizin çok üstünde. Kanser ve kalp hastalıklarının teşhisinde –röntgenlerin okunmasında bile– uzmanlar görüş ayrılıkları yaşayabiliyorlar. Bu, bir hastanın aldığı tedavinin bir piyangonun sonucu olabileceği anlamına geliyor."

"Doktorlar pazartesi veya cuma günü, sabahın erken saati ya da akşamüstü hep aynı kararı verdiklerine düşünmeyi tercih ederler. Ancak öyle görünüyor ki doktorların söyledikleri ve yaptıkları ne kadar yorgun olduklarına göre değişebiliyor."

"Tıbbi kılavuzlar doktorların hastanın aleyhine gaf yapma olasılığını azaltabilir. Bu tür kılavuzlar ayrıca tıp mesleğine bütün olarak fayda sağlayabilir, çünkü değişkenliği azaltırlar."

BÖLÜM 23

Performans Puanlamalarında Ölçeği Tanımlamak

Bir egzersizle başlayalım. Tanıdığınız üç kişiyi alın; arkadaş ya da meslektaş olabilirler. Onları, iyilik, zekâ ve çalışkanlık özellikleri açısından 1'in en düşük, 5'in en yüksek puan olduğu 1 - 5 ölçeğinde puanlayın. Şimdi onları iyi tanıyan birinden –eşiniz, en yakın arkadaşınız ya da en yakın çalışma arkadaşınız– aynı üç insan için, aynı şeyi yapmasını rica edin.

Puanlamaların bazılarında, sizin ve diğer değerlendiricinin farklı sayılar vermiş olmanız ihtimali çok yüksek. Siz (ve diğer değerlendirici) buna istekliyseniz, lütfen farklılığın nedenlerini tartışın. Cevabın, yatay gürültü olarak adlandırdığımız, ölçeği kullanma şeklinizde yattığını görebilirsiniz. Belki siz 5 puanın gerçekten olağanüstü iyi bir şeyi gerektirdiğini düşündünüz, diğer değerlendirici sadece normalin üstünde iyi bir şeyin yeterli olduğunu düşündü. Ya da belki de puanlanan kişiyle ilgili farklı görüşleriniz yüzünden farklı puanlar verdiniz; kişinin iyi olup olmaması konusundaki anlayışınız ve bu erdemi tanımlama şekliniz diğer değerlendiriciden farklı olabilir.

Şimdi puanladığınız üç kişi için ortada bir terfi ya da bonus olduğunu düşünün. Siz ve diğer değerlendirici iyiliğe (ya da meslektaş dayanışmasına), zekâya ve çalışkanlığa değer veren bir şirkette performans puanlama işinden sorumlusunuz. Puanlarınızın arasında bir fark olur muydu? Fark bir önceki egzersizdeki kadar büyük olur muydu? Yoksa daha mı büyük olurdu? Bu sorular her nasıl cevaplanırsa cevaplansın, politikalar ve ölçeklendirmedeki farkların gürültü yaratması olasıdır. Ve aslında örgütsel ortamlarda performans puanlamalarında yaygın olarak gözlenen şey budur.

Bir Yargı Görevi

Neredeyse bütün büyük örgütlerde, performans düzenli olarak resmî şekilde değerlendirilir. Puanlananlar bu tecrübeden keyif almazlar. Bir gazete manşetinin ifade ettiği gibi, "Bir Çalışma Neredeyse Herkesin Performans Değerlendirmelerinden Nefret Ettiğini Ortaya Koydu." Söz konusu herkes ayrıca, performans değerlendirmelerinin hem ön yargıya hem gürültüye açık olduğunu da bilir (bizce). Ama çoğu insan, ne kadar gürültülü olduklarından habersizdir.

İdeal bir dünyada, insanların performansını değerlendirmek bir yargı görevi olmazdı: Nesnel olgular insanların ne kadar iyi iş çıkardığını saptamaya yeterdi. Ancak çoğu modern örgütün Adam Smith'in her işçinin ölçülebilir bir çıktısının olduğu iğne fabrikasıyla çok az ortak noktası var. Bu çıktı, bir CFO ve araştırma şefi için ne olurdu? Bugünün bilgi işçileri, birden fazla, hatta bazen birbiriyle çelişen hedefleri dengeliyorlar. O hedeflerden sadece birine odaklanmak hatalı değerlendirmeler üretebilir ve zarar verici teşvik etkileri yaratabilir. Bir doktorun her gün gördüğü hasta sayısı örneğin hastanenin üretkenliği açısından önemli bir yönlendiricidir ama doktorların sadece bu göstergeye odaklanmasını, hele sadece bu temele göre değerlendirilip ödüllen-

dirilmesini istemezsiniz. Sayıya dökülebilen performans ölçümleri bile –mesela bir satışçı için satış miktarı ve bir programcı için yazılan kod satırı sayısı gibi– bağlama göre değerlendirilmelidir: Bütün müşterilere hizmet etmek eşit derecede zor değildir ve bütün yazılım geliştirme projeleri aynı değildir. Bu zorlukların ışığında, pek çok insan tamamen nesnel performans ölçülerine göre değerlendirilemez. Yargıya dayalı performans değerlendirmelerine her yerde rastlanması bu yüzdendir.

Dörtte Bir Sinyal, Dörtte Üç Gürültü

Performans değerlendirmelerinin pratiği üzerine binlerce araştırma makalesi yayımlandı. Çoğu araştırmacı bu tür değerlendirmelerin son derece gürültülü olduğu sonucuna vardı. Bu sarsıcı sonuç büyük ölçüde, birden fazla değerlendiricinin aynı kişi için genellikle performansın birçok boyutu üzerinden girdi sağladığı 360 derece performans değerlendirmelerine dayanıyor. Bu analizin yapıldığı yerlerde sonuç çok sevimli olmaz. Çalışmalar genellikle gerçek farklılığın, yani kişinin performansına atfedilebilecek farklılığın, toplam farklılığın en %20 ve %30'una karşılık geldiğini gösteriyor. Geri kalanı, yani puanlamadaki farklılığın %70 ila %80'i sistem gürültüsünden kaynaklanıyor.

Bu gürültü nereden geliyor? İş performansı puanlamalarındaki farklılıklarla ilgili sayısız çalışma sayesinde, bu tür değerlendirmelerde sistem gürültüsünün bütün ögelerinin olduğunu biliyoruz.

Bu ögeleri bir performans puanlaması bağlamında gözümüzde daha kolay canlandırabiliriz. Lynn ve Mary adında iki değerlendiricimiz olsun. Lynn'in ortalamada Mary'den daha yüksek puanlar vermesi anlamında, Lynn daha merhametli ve Mary daha sert ise, burada yatay gürültü söz konusudur. Karar vericilerle ilgili tartışmamızda değindiğimiz gibi, bu gürültü Lynn ve Mary'nin ya gerçekten farklı izlenimlere sahip olduğu ya da iki değerlendiricinin

aynı izlenimi ifade ederken puanlama ölçeğini farklı kullandıkları anlamına gelebilir.

Şimdi, eğer Lynn sizi değerlendiriyorsa ve sizi ve katkılarınızı bariz bir şekilde yetersiz buluyorsa, geneldeki yumuşaklığı size olan kişiye özel (ve olumsuz) tepkisiyle dengelenebilir. Buna sabit örüntü diyoruz: belli bir değerlendiricinin belli bir kişiye tepkisinin puanlanması. Örüntü Lynn'e (ve sizinle ilgili yargısına) has olduğu için, bir örüntü gürültüsü kaynağıdır.

Son olarak, Mary bir puanlama formunu doldurmadan hemen önce şirket otoparkında arabasının çizildiğini keşfetmiş olabilir ya da biraz önce kendi şaşırtıcı derecede cömert primini alması onu sizin performansınızı değerlendirirken her zamankinden iyi bir ruh hâline sokmuş olabilir. Elbette bu tür olaylar durum gürültüsü yaratabilir.

Farklı çalışmalar sistem gürültüsünün bu üç ögeye (seviye, örüntü ve durum) ayrılması konusunda farklı sonuçlara ulaştılar ve elbette, sonuçların bir örgütten diğerine değişebilmesinin nedenlerini tahmin edebiliyoruz. Ancak bütün gürültü biçimleri sevimsizdir. Bu araştırmadan çıkan temel mesaj çok basit: Çoğu performans değerlendirmesi, puanlanan kişinin kendisiyle arzu edeceğimizden daha az alakalıdır. Bir çalışmanın özetlediği gibi, "İş performansı ve iş performansının puanlanması arasındaki ilişkinin zayıf ve en iyi ihtimalle belirsiz olması çok mümkündür."

Ek olarak, örgütlerdeki puanlamaların değerlendiricinin çalışanın gerçek performansıyla ilgili algısını yansıtmamasının birçok nedeni olabilir. Örneğin, değerlendirici performansı doğru bir şekilde değerlendirmek yerine, insanları "stratejik" olarak puanlayabilir. Değerlendiriciler, başka dürtülerin yanı sıra, zor bir geri besleme konuşmasından kaçınmak, uzun süredir terfi bekleyen bir insanı kayırmak ve hatta paradoksal bir şekilde, bir başka bölüme transfer olmak için iyi bir değerlendirmeye ihtiyacı olan bir takım üyesinden kurtulmak için puanlamayı kasten şişirebilirler.

Bu stratejik hesaplamalar elbette puanlamaları etkiler ama tek gürültü kaynağı onlar değildir. Bunu bir tür doğal test sayesinde biliyoruz. Bazı 360 derece geri besleme sistemleri sadece gelişimsel amaçlar için kullanılır. Bu sistemlerde, cevap veren kişilere geri beslemenin değerlendirme amacı için kullanılmayacağı söylenir. Değerlendiricilerin onlara söylenenlere inanması ölçüsünde, bu yaklaşım onları puanlamaları şişirmekten –ya da söndürmekten– alıkoyabilir. Sonuç olarak görülür ki gelişimsel değerlendirme geri beslemenin kalitesinde bir fark yaratmaktadır yaratmasına ama sistem gürültüsü hâlâ yüksektir ve farklılıktan, puanlanan kişinin performansından hâlâ fazla sorumludur. Geri besleme sadece gelişim amaçlı olduğunda bile, puanlamalar gürültülü olmaya devam eder.

Uzun Zamandır Bilinen ama Çözülmeyen bir Sorun

Performans puanlama sistemleri bu kadar bozuksa, performans ölçen insanların bunu dikkate alıp sistemleri iyileştirmesi gerekir. Hatta, son birkaç on yılda, örgütler bu sistemlerde sayısız reform denemesi yaptılar. Bu reformlarda değindiğimiz gürültü azaltma stratejilerinin bazıları kullanıldı. Bize göre, çok daha fazlası yapılabilirdi.

Hemen hemen bütün örgütler gürültü azaltma stratejisi olarak *bir araya toplama*yı kullanıyorlar. Bir araya toplanan puanlamalar genellikle 1990'larda büyük şirketlerde standartlaşan 360 derece puanlama sistemleriyle ilişkilendirilir. (*Human Resources Management* dergisi, 1993'te 360 derece geri besleme üzerine özel bir sayı yayımladı.)

Çok sayıda değerlendiricinin puanlamalarının ortalamasını almak sistem gürültüsünü azaltmaya yardımcı olsa da, 360 derece geri besleme sistemlerinin bu soruna çare olarak icat edilmediğini belirtmekte fayda var. Bu sistemlerin başlıca amacı bir

patronun gördüğünden fazlasını ölçebilmektir. Sadece üstleriniz değil, eş pozisyonda olduğunuz kişiler ve altınızdaki personelin de performans değerlendirmenize katkıda bulunmaları istendiğinde, değer verilen şeyin doğası değişir. Teori bu değişikliğin iyi yönde olduğunu söyler; çünkü bugünün meslekleri patronu ya da üstünüzü hoşnut etmekten daha fazlasını gerektirir. 360 derece geri beslemenin popülerliğindeki artış, akışkan ve proje bazlı örgütlerin yaygınlaşmasıyla aynı zamana denk geldi.

360 derece geri beslemenin nesnel olarak ölçülebilir performansı öngörmesi açısından faydalı bir araç olduğunu gösteren birtakım kanıtlar mevcut. Ne yazık ki bu geri besleme sisteminin kullanımı kendi sorunlarını da yarattı. Otomasyon geri besleme sistemlerine soru eklemeyi daha zahmetsiz kıldığı ve çoklu kurumsal hedeflerin ve kısıtlamaların çoğalması iş tanımlarına yeni boyutlar eklediği için, pek çok geri besleme anketi saçmalık derecesinde karmaşık bir hâl aldı. Gereğinden fazla çalışılmış ve üstünde oynanmış anketler o kadar çok ki (her değerlendirici ve puanlanan kişi için on bir boyutta kırk altı puanlama olan örnekler var). Bu kadar çok farklı boyutta değerlendirilen çok sayıda insanın her biri için doğru ve ilgili olguları hatırlamak ve süzgeçten geçirmek, insanüstü bir değerlendirici olmayı gerektirir. Bazı açılardan bu aşırı karmaşık yaklaşım sadece işe yaramaz değil, aynı zamanda zararlıdır. Daha önce gördüğümüz gibi, halo etkisi sözde ayrı boyutların aslında ayrı ayrı değerlendirilmeyeceğini düşündürür. İlk sorulardan birine verilen güçlü bir olumlu ya da olumsuz puan, sonraki soruların cevaplarını aynı yöne çekecektir.

Daha önemlisi 360 derece sistemlerinin geliştirilmesi geri besleme sağlamaya ayrılan zamanı katlayarak artırdı. Orta düzey yöneticilerden her seviyedeki çalışma arkadaşları ve bazen başka örgütlerdeki mevkidaşları için düzinelerce anket doldurmasının istenmesi artık nadir bir durum olmaktan çıktı, çünkü artık birçok şirket müşterilerinden, satıcılarından ve diğer iş ortaklarından da

geri besleme talep ediyor. Niyet ne kadar iyi olursa olsun, zamanı kısıtlı olan değerlendiricilere yüklenen bu talep patlamasının, bu kişilerin sağladığı bilginin kalitesini iyileştirmesi beklenemez. Bu durumda, gürültünün azaltılması ortaya çıkan maliyete değmeyebilir. Bu sorunu Kısım 6'da tartışacağız.

Son olarak, 360 Derece Sistemleri bütün performans ölçüm sistemlerinin neredeyse evrensel hastalığına, gittikçe artan puanlama enflasyonuna bağışık değildir. Büyük bir endüstri şirketi bir keresinde, yöneticilerinin %98'inin "beklentileri tam olarak karşılar" şeklinde puanlandığını saptadı. Neredeyse herkes mümkün olan en yüksek puanı alınca, bu puanlamaların değerini sorgulamak kaçınılmaz olur.

Göreceli Yargılara Övgü

Puanlama enflasyonuna teoride etkili bir çözüm, puanlamaya bir tür standardizasyon katmaktır. Bunu yapmayı amaçlayan popüler bir uygulama *zorunlu sıralamadır.* Zorunlu sıralama sisteminde, sadece değerlendiricilerin herkese en yüksek olası puanı vermesi önlenmiş olmaz, aynı zamanda önceden belirlenmiş bir dağılıma uymaya da mecbur tutulurlar. Zorunlu sıralama, General Electric'in CEO'su olduğu dönemde Jack Welch tarafından puanlamalardaki enflasyonu durdurmanın ve performans değerlendirmelerinde "samimiyeti" sağlamanın bir yolu olarak savunulmuştu. Pek çok şirket bu uygulamayı benimsedi ancak daha sonra moral ve takım çalışması üstündeki nahoş etkilerini sıralayarak vazgeçtiler.

Kusurları her ne olursa olsun, sıralamalar puanlamalardan daha az gürültülüdür. Ceza tazminatları örneğinde, göreceli yargılarda mutlak yargılarda olduğundan çok daha az gürültü olduğunu görmüştük. Aynı ilişkinin performans değerlendirmelerinde de geçerli olduğu saptandı.

Panel A

A Çalışanının iş kalitesi:

1	2	3	4	5
Çok kötü	Kötü	Makul	İyi	Mükemmel

Panel B

Astlarınızı *güvenlik* konusunda puanlayın. *Güvenlik* çalışanların gerekli kural ve düzenlemelere ne kadar uyduğunu, işte güvenli bir şekilde hareket edip etmediğini ve güvenli çalışma uygulamaları konusunda farkındalık ve anlayış sahibi olup olmadığını içerir.



ŞEKİL 17: *Mutlak ve göreceli puanlama ölçeklerine örnekler*

Nedenini kavrayabilmek için, çalışan değerlendirme ölçeklerinden iki örneği gösteren Şekil 17'ye bakalım. Bir çalışanın mutlak ölçek üzerinde değerlendirildiği Panel A, eşleme operasyonu dediğimiz şeyi gerektirir: Çalışanın "iş kalitesi" konusundaki izleniminizle en yakın eşleşen puanı bulmanız gerekir. Panel B ise, aksine, her bireyin belirli bir boyutta (güvenlik) bir grup insanla karşılaştırılmasını gerektirir. Yöneticiden bir persentil ölçeği kullanarak, çalışanın belirli bir nüfus içindeki sıralamasını belirtmesi istenir. Yöneticinin üç çalışanı bu ortak ölçeğe yerleştirdiğini görüyoruz.

Panel B'deki yaklaşımın iki avantajı var. Öncelikle, bütün çalışanları tek seferde bir boyut üzerinde (bu örnekte güvenlik) puanlamak, bir sonraki bölümde detaylı olarak ele alacağımız gürültü azaltma stratejisine –karmaşık bir yargıyı birden fazla boyutta *yapılandırma*– iyi bir örnektir. Yapılandırma halo etkisini azaltmayı amaçlar ve bu, normalde bir bireyin farklı boyutlardaki puanla-

malarını küçük bir aralıkta tutmaya yarar. (Elbette yapılandırma sadece, bu örnekte olduğu gibi, sıralandırma her boyutta ayrı ayrı yapılırsa işe yarar. Çalışanların, yetersiz tanımlanmış, toplama bir "iş kalitesi" yargısına göre sıralanması halo etkisini azaltmaz.)

İkinci olarak Bölüm 15'te değindiğimiz gibi, sıraya koymak hem örüntü gürültüsünü hem yatay gürültüyü azaltır. Takımınızın iki üyesinin performanslarını karşılaştırdığınız zaman, ikisine ayrı ayrı puan verirken olduğunuzdan daha tutarlı olmanız beklenir. Daha önemlisi, sıralama yapmak yatay gürültüyü mekanik olarak ortadan kaldırır. Lynn ve Mary yirmi çalışandan oluşan aynı grubu değerlendiriyorsa ve Lynn Mary'den daha merhametliyse, ortalama puanları farklı olur ama ortalama sıralamaları farklı olmaz. Merhametli bir değerlendirici ve sert bir değerlendirici aynı sıralamaları kullanır.

Aslında, bütün değerlendiricilerin aynı ortalamaya ve aynı değerlendirme dağılımına sahip olmasını sağlayan zorunlu sıralamanın başlıca hedefi gürültünün azaltılmasıdır. Puan dağılımı mecbur tutulduğunda sıralama "zorunlu" olur. Örneğin bir kural puanlanan insanların en çok %20'sinin en üst kategoriye yerleştirilebileceğini ve en az %15'inin en alt kategoriye yerleştirilmesi gerektiğini belirtebilir.

Sırala ama Zorlama

Dolayısıyla prensipte, zorunlu sıralamanın çok ihtiyaç duyulan iyileştirmeleri sağlaması gerekir. Ancak bu yöntem sıklıkla geri teper. Burada bütün olası istenmeyen etkileri (ki genellikle prensipten çok yetersiz uygulamayla ilişkilidirler) gözden geçirmek niyetinde değiliz. Ancak zorunlu sıralama sistemine dair iki mesele bazı genel dersleri içermektedir.

İlki mutlak ve göreceli performans arasındaki karmaşadır. Bir şirketin yöneticilerinin %98'inin emsallerinin arasında en üst

%20'lik, %50'lik hatta %80'lik diliminde olması imkânsızdır. Ancak söz konusu beklentiler öncesinde ve *mutlak şekilde* tanımlanmışsa, hepsinin "beklentileri karşılaması" imkânsız değildir.

Pek çok üst düzey yönetici neredeyse bütün çalışanların beklentileri karşılayabildiği kavramına itiraz eder. Öyle olması için, belki de hoşnutluk kültürü nedeniyle, beklentilerin çok düşük olması gerektiğini savunurlar. Bu yorumun geçerli olabileceğini kabul etmeliyiz ancak çoğu çalışanın *yüksek* beklentileri gerçekten karşılıyor olması da mümkündür. Hatta yüksek performanslı bir örgütte bulmayı beklediğimiz budur. Başarılı bir uzay görevindeki bütün astronotların beklentileri tam anlamıyla karşıladığını duysanız, Ulusal Havacılık ve Uzay Dairesi'nin performans yönetimlerinin fazla müsamahalı olduğunu düşünüp dudak bükmezsiniz.

Kısacası, göreceli değerlendirmelere dayanan bir sistem sadece örgüt göreceli performansı önemserse uygundur. Örneğin, göreceli puanlamalar, kişilerin mutlak performansından bağımsız olarak, sadece sabit bir yüzdelerinin terfi alabilmesi durumunda anlamlı olabilir. Generalliğe terfi ettirilen albayları düşünün. Ancak pek çok şirketin yaptığı gibi, *mutlak* performans seviyesini ölçme iddiasını taşıyan bir sisteme göreceli bir sıralama dayatmak, mantıksızdır. Ve belirli bir çalışan yüzdesinin (mutlak) beklentileri karşılayamıyor olarak puanlanmasını zorunlu kılmak da sadece acımasızlık değildir, aynı zamanda saçmadır. Ordunun elit bir biriminin %10'unun "yetersiz" olarak sınıflandırılması gerektiğini söylemek aptalcadır.

İkinci sorun, puanlamaların zorunlu dağılımının altta yatan gerçek performansları –tipik olarak normal dağılıma yakın– yansıttığının varsayılmasıdır. Oysa puanlanan nüfustaki performansların dağılımı biliniyor bile olsa, aynı dağılım, örneğin tek bir değerlendirici tarafından değerlendirilenler gibi, daha küçük bir grupta yeniden üretilmeyebilir. Binlerce kişilik bir nüfustan rastgele on kişi seçerseniz, tam olarak ikisinin genel nüfusun en üst

%20'lik dilimine ait olduğu garanti değildir. (Garanti değildir hafif bir ifade; bunun gerçekleşme olasılığı sadece %30'dur.) Pratikte, sorun daha da kötüdür çünkü takımların kompozisyonu rastgele değildir. Bazı birimlerinin neredeyse tamamı yüksek performans sahibi çalışanlardan, diğerleri ise ortalamanın altındaki çalışanlardan oluşuyor olabilir.

Kaçınılmaz olarak, böyle bir ortamda zorunlu sıralama bir hata ve haksızlık kaynağı olur. Bir değerlendiricinin takımının performansları birbirinden ayırt edilemez beş kişiden oluştuğunu varsayın. Bu farklılaşmamış gerçekliğe farklılaşmış bir puan dağılımını dayatmak hatayı azaltmaz. Artırır.

Zorunlu sıralamayı eleştirenler saldırılarını genellikle gaddarca, insanlık dışı ve nihayetinde üretkenlik karşıtı buldukları sıralama prensibine odaklarlar. Bu argümanları kabul edin ya da etmeyin, zorunlu sıralamanın ölümcül kusuru "sıralama" değil, "zorunlu" olmasıdır. Yargılar uygun olmayan bir ölçeğe zorla dayatıldığında, sebebi ister mutlak bir performansı ölçmek için göreceli bir ölçeğin kullanılması olsun, ister karar vericilerin ayırt edilemezi ayırt etmeye zorlanmaları olsun, ölçek tercihi mekanik olarak gürültü katacaktır.

Sırada Ne Var?

Örgütlerin performans ölçümünü iyileştirmek için harcadığı bütün çabaların ışığında, sonuçların hayal kırıklığı yarattığını söylemek hafif kalır. O çabaların sonucu olarak, performans değerlendirmenin maliyeti tırmanışa geçti. 2015 yılında, Deloitte altmış beş bin çalışanını değerlendirmeye her yıl 2 milyon saat harcadığını hesapladı. Performans görüşmeleri örgütlerin en nahoş bulunan ritüelleri arasında yer almaya devam ediyor ve görüşmelerin iki tarafında da eşit derecede nefret uyandırıyor. Bir çalışma, yöneticilerin, çalışanların ve İK yöneticilerinin %90 gibi çarpıcı bir

oranının performans yönetimi süreçlerinin bekledikleri sonuçları vermediğine inandıklarını ortaya koydu. Araştırma birçok yöneticinin yaşadığı şeyi teyit etti. Her ne kadar performans geri beslemesi çalışan için bir gelişim planıyla ilişkilendirildiğinde iyileşmeyi getirebilse de, performans puanlamaları şu anda uygulandıkları hâlleriyle motive ettikleri kadar motivasyonu da bozabiliyorlar. Bir inceleme makalesinin özetlediği gibi, "Onlarca yıl içinde [performans yönetimi] süreçlerini iyileştirmek için denenen onca şeye rağmen, yanlış bilgi üretmeye devam ediyor ve performansı artırmak için neredeyse hiçbir şey yapmıyorlar."

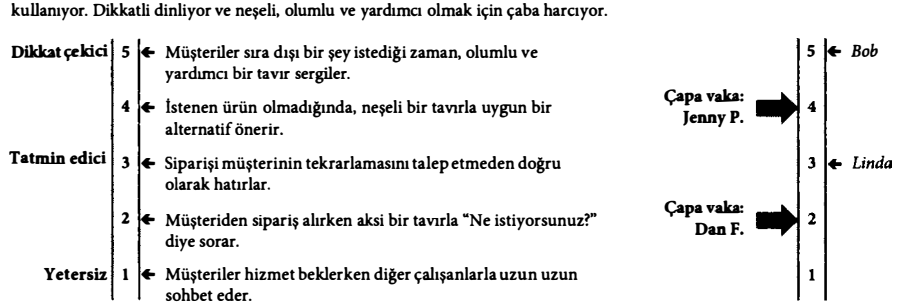
Çaresizlik, şimdilik az sayıda olmakla birlikte sayıları her geçen gün artan bazı şirketlere, değerlendirme sistemini tamamen kaldırmak gibi radikal bir seçeneği düşündürüyor. Bu "performans yönetimi devrimi"nin aralarında birçok teknoloji şirketinin, profesyonel hizmet örgütlerinin ve geleneksel sektörlerdeki bir avuç şirketin de bulunduğu savunucuları, geriye dönük bu değerlendirme yerine, gelişim ve gelecek odaklı geri beslemeye odaklanmayı amaçlıyorlar. Değerlendirmelerinden sayıları kaldırarak geleneksel performans puanlamalarını terk edenler de var.

Performans puanlamalarından vazgeçmeyen şirketler için (ki hâlâ ezici çoğunluğu oluşturuyorlar), iyileşme yolunda neler yapılabilir? Bir kez daha, gürültüyü azaltma stratejilerinden biri, doğru ölçeği seçmekle başlıyor. Amaç *ortak bir referans çerçevesi* sağlamak. Araştırmalar, iyileştirilen puanlama formatlarının ve değerlendiricilerin eğitilmesinin, değerlendiriciler arasında ölçek kullanımında daha fazla tutarlılığın sağlanmasına yardımcı olabileceğini gösteriyor.

Performans puanlama ölçekleri, asgari düzeyde, tutarlı olarak yorumlanmak için yeterince özellikli tanımlayıcılara dayandırılmalıdır. Pek çok örgüt ölçekteki her derecenin belirli bir davranışın tanımına karşılık geldiği *davranışa dayalı değerlendirme ölçekleri* kullanır. Şekil 18'in sol paneli buna bir örnektir.

Öte yandan, kanıtlar davranışa dayalı değerlendirme ölçeklerinin gürültüyü ortadan kaldırmaya yetmediğini düşündürmektedir. Daha ileri bir adım olarak, *referans çerçevesi eğitiminin* değerlendiricilerin arasında tutarlılığın sağlanmasına yardımcı olduğu görülmüştür. Bu adımda, değerlendiriciler farklı performans boyutlarını ayırt etmek için eğitilirler. Videoya kaydedilmiş hikâyeler kullanılarak performans puanlama egzersizleri yapar ve kendi puanlamalarını uzmanlar tarafından sağlanan "gerçek" puanlamalarla karşılaştırabilirler. Performans hikâyeleri referans vaka görevi üstlenir ve her biri performans ölçeğinde bir çapa noktası sunar. Böylece ölçek, Şekil 18'in sağ panelinde gösterildiği gibi bir *vaka ölçeğine* dönüşür.

Bir vaka ölçeğinde, yeni bir bireyin her puanlaması çapa vakalarla bir karşılaştırmadır. Göreceli bir yargıya dönüşür. Karşılaştırmalı yargılar gürültüye puanlamalardan daha az açık olduğu için, vaka ölçekleri sayı, sıfat ya da davranışsal tanımlamalar kullanan ölçeklerden daha güvenilirdir.



ŞEKİL 18: *Davranışa dayalı değerlendirme ölçeği örneği (sol) ve vaka ölçeği örneği (sağ)*

Referans çerçevesi eğitimi onlarca yıldır bilinmektedir ve kanıtlanabilir şekilde çok daha az gürültü ve daha isabetli pu-

anlamalar sağlamaktadır. Buna rağmen çok rağbet görmemiştir. Nedenini tahmin etmek zor değil. Referans çerçevesi eğitimi, vaka ölçekleri ve aynı amaçları güden diğer araçlar karmaşık ve zaman alıcıdır. Değerli olmak için genellikle şirkete, hatta değerlendirmeleri yürüten birime göre uyarlanmaları ve görevlerin gereklilikleri geliştikçe onların da sık sık güncellenmesi gerekir. Bu araçlar şirketin performans yönetimi sistemlerinde zaten büyük olan yatırımlarına ek yapmasını gerektirir. Oysa şu anda moda tam aksi yöne kayıyor (Kısım 6'da gürültüyü azaltmanın maliyetleri konusuna daha çok değineceğiz).

Ek olarak değerlendiricilere atfedilebilecek gürültüyü ehlileştiren her örgüt aynı zamanda puanlamaları kendi amaçlarının doğrultusunda etkileme becerisini de azaltır. Yöneticilerden ek değerlendirici eğitimleri almalarını, puanlama sürecine daha fazla çaba harcamalarını ve sonuçlar üstündeki kontrollerinin bir kısmından vazgeçmelerini talep etmenin bir miktar direnişe neden olacağı kesindir. Referans çerçevesi değerlendirici eğitimleriyle ilgili çalışmaların çoğunluğunun gerçek yöneticiler değil, öğrenciler üstünde gerçekleştirilmesi çok şey anlatıyor.

Performans değerlendirmesi konusu hem pratik hem felsefi açıdan birçok soruyu akla getiriyor. Örneğin bazı insanlar bugünün sonuçların genellikle insanların birbirleriyle etkileşimlerine bağlı olduğu örgütlerinde bireysel performans kavramının ne derece anlamlı olduğunu soruyorlar. Kavramın anlamlı olduğuna inanıyorsak, bireysel performans seviyelerinin bir örgütteki insanlar arasında nasıl dağıtıldığını –örneğin, performans normal bir dağılım mı izliyor yoksa son derece orantısız bir katkı sağlayan bir "yıldız yetenek" mi var?– sorgulamalıyız. Ve eğer amacınız insanların en iyi taraflarını ortaya çıkarmaksa, akla yatkın bir şekilde, bireysel performansı ölçmenin ve o ölçümü insanları korku ve açgözlülükle motive etmede kullanmanın en iyi yaklaşım (hatta etkili bir yaklaşım) olup olmadığını sorabilirsiniz.

Bir performans yönetimi sistemi tasarlıyor ya da yeniliyorsanız, bu soruları ve daha pek çoğunu cevaplamanız gerekir. Burada gayemiz bu soruları incelemek değil, daha mütevazı bir öneride bulunmak: Performans ölçüyorsanız, performans puanlamalarınız muhtemelen sistem gürültüsünün istilasına uğramıştır ve bu nedenle, işe yaramaz ve hatta amaca olumsuz etkili olmaları mümkündür. Bu gürültüyü azaltmak basit teknolojik önlemlerle çözülebilecek bir sorun değildir. Değerlendiricilerin yapmalarının beklendiği yargılar hakkında net düşünmeyi gerektirir. Muhtemelen, puanlama ölçeğini netleştirerek ve insanları onu tutarlı olarak kullanabilmeleri için eğiterek, yargıları iyileştirebileceğinizi göreceksiniz. Bu gürültü azaltma stratejisi pek çok diğer alanda uygulanabilir.

Ölçüyü Tanımlamaktan Bahsederken

"Performans puanlamalarımızda çok zaman harcarız ama yine de sonuçlar dörtte bir performans, dörtte üç sistem gürültüsünden oluşur."

"Bu sorunu gidermek için 360 derece geri beslemeyi ve zorunlu sıralamayı denedik ama işleri daha da kötüleştirmiş olabiliriz."

"Çok fazla yatay gürültü varsa, nedeni farklı değerlendiricilerin, "iyi" ve "harika"nın ne anlama geldiği konusunda bambaşka fikirlerinin olmasıdır. Sadece onlara puanlama ölçeğinde çapa olarak somut vakalar sağlarsak hemfikir olabilirler."

BÖLÜM 24

İşe Alımda Yapılandırma

Herhangi bir işte çalıştıysanız, iş mülakatı kelimeleri size çok canlı ve stresli anılarınızı çağrıştırıyor olabilir. Bir adayın gelecekteki amiriyle ya da bir İK profesyoneliyle bir araya geldiği iş mülakatları pek çok örgüte giriş için gerekli görülen bir geçiş ayinidir.

Pek çok örnekte, mülakatlar iyi prova edilmiş bir rutine uygun ilerler. Birkaç espri alışverişinden sonra, mülakatı yapan kişi adaydan tecrübesini tanımlamasını ve belirli yönlerini detaylı olarak anlatmasını rica eder. Başarılar ve mücadelelerle, iş konusunda motivasyonla ilgili sorular ya da şirket için iyileştirici fikirler sorulur. Mülakatçı sıklıkla adaydan kişiliğini tarif etmesini ve pozisyona ya da şirketin kültürüne neden uygun olduğunu anlatmasını ister. Hobiler ve ilgi alanları da zaman zaman tartışılır. Sona doğru, genellikle adaya da birkaç soru sorma fırsatı verilir ve bu sorular alaka ve içgörü açısından usulünce değerlendirilir.

Şu anda eleman alacak bir pozisyondaysanız, seçme yöntemleriniz muhtemelen bu ritüelin bir versiyonunu içeriyordur. Bir örgüt psikoloğunun ifade ettiği gibi, "Bir kişinin mülakat yapılmadan işe alınması nadir, hatta düşünülemez bir durumdur." Ve neredeyse

bütün profesyoneller bu mülakatlarda işe alma kararı verirken, bir dereceye kadar sezgisel yargılarına güvenirler.

İşe alım mülakatının çok yaygın olması, konu birlikte çalışacağımız insanları seçmek olduğunda, yargının değerine olan derine işlemiş inancımızı yansıtır. Ve personel seçiminin bir yargı görevi olarak büyük bir avantajı vardır: Çok yaygın ve önemli olduğu için, örgüt psikologları bu konuyu detaylı olarak incelediler. *Journal of Applied Psychology*'nin 1917'de yayımlanan çıkış sayısı, işe alımı "en büyük sorun..." olarak tanımlar, "... çünkü ne de olsa insan becerileri başlıca ulusal kaynaktır." Bu makaleden bir yüzyıl sonra, artık çeşitli seçme tekniklerinin (standart mülakatlar dâhil) etkinliği hakkında çok şey biliyoruz. Hiçbir karmaşık yargı görevi bu kadar çok saha araştırmasının odak noktası olmamıştır. Bu durum, işe alımları, çok sayıda seçenek arasından seçim yapmayı gerektiren pek çok yargı için anlam çıkarılabilecek dersler sunan kusursuz bir test örneği yapar.

Mülakatların Tehlikeleri

İşe alım mülakatlarıyla ilgili araştırmalara yabancıysanız, bundan sonra okuyacaklarınız sizi şaşırtabilir. Kısaca, amacınız bir işte hangi adayların başarılı olacağını, hangilerinin başarısız olacağını saptamaksa standart mülakatlar (birazdan değineceğimiz yapılandırılmış mülakatlardan ayırmak için bu mülakat türüne yapılandırılmamış mülakatlar da diyoruz) çok bilgilendirici olmayacaktır. Daha açık ifade etmek gerekirse, genellikle bir işe yaramazlar.

Bu sonuca ulaşmak için, sayısız çalışmada bir değerlendiricinin mülakat sonrası bir adaya verdiği puanla aynı adayın ilerleyen zaman içinde o işteki başarısı arasındaki korelasyon tahmin edildi. Mülakat puanıyla başarı arasındaki korelasyon yüksekse, mülakatlar –ya da korelasyonun aynı şekilde hesaplandığı diğer işe alım

teknikleri– adayların nasıl performans sergileyeceği konusunda iyi bir kestirici olarak kabul edilebilirdi.

Burada bir ikaza gereksinim var. Başarının tanımı hafife alınmayacak bir sorundur. Tipik olarak, performans kişinin üstünün (amirinin) puanlamaları dayanak alınarak değerlendirilir. Bazen ölçü istihdam süresidir. Bu tür ölçüler elbette birtakım soruları doğurur, özellikle bir önceki bölümde değindiğimiz performans değerlendirmelerinin geçerliliği sorunu göz önüne alındığında. Ancak, bir işverenin personel seçerken ortaya koyduğu yargılarının kalitesini değerlendirme amacı açısından, aynı işverenin işe alınmış personeli değerlendirirken vardığı yargıları kullanmak akla yatkın görünmektedir. İşe alım kararlarının kalitesi üzerindeki her analiz bu varsayımdan yola çıkmalıdır.

Öyleyse, bu analizlerin vardığı sonuç nedir? Bölüm 11'de tipik mülakat puanlamaları ve iş performans puanlamaları arasında 0,28'lik bir korelasyondan bahsetmiştik. Başka çalışmalar 0,20 ile 0,33 arasında değişen korelasyonlar bildiriyorlar. Daha önce gördüğümüz gibi, bu, sosyal bilim standartları açısından çok iyi bir korelasyon olmakla birlikte, kararlarınızı dayandırabileceğiniz kadar iyi bir korelasyon değil. Bölüm 3'te değindiğimiz yüzdelik uyumu (YU) kullanarak bir olasılık hesaplayabiliriz: Yukarıdaki korelasyon seviyelerinde, iki aday hakkında bütün bildiğiniz birinin mülakatta diğerinden daha iyi göründüğüyse, daha iyi görünen bu adayın daha iyi performans sergileme olasılığı %56 ile %61 arasında değişecektir. Bir şekilde yazı tura atmaktan daha iyi olduğu kesin ama önemli kararlar almak için çok emin bir yol olduğu söylenemez.

Mülakatların bir aday hakkında yargıya varmanın yanı sıra başka amaçlara da hizmet ettiği doğrudur. Şirketi gelecek vadeden adaylara satmak ve müstakbel çalışma arkadaşlarıyla bir ilişki inşa etmeye başlamak için bir fırsat sundukları kesindir. Ancak yetenek seçimine zaman ve çaba harcayan bir örgütün bakış açısıyla,

mülakatların ana amacı açıkça seçme yapmaktır. Ve bu görevde müthiş bir başarı sağladıkları söylenemez.

Mülakatta Gürültü

Geleneksel mülakatların iş performansı tahmininde neden hata ürettiklerini kolayca anlayabiliriz. Bu hata kısmen objektif bilgisizlik olarak adlandırdığımız (Bölüm 11) şeyle ilgilidir. İş performansı, işe aldığınız kişinin yeni pozisyonuna ne kadar hızlı uyum sağladığı ve çeşitli hayat olaylarının çalışmasını nasıl etkilediği gibi birçok şeye bağlıdır. Bunların büyük kısmı işe alım sırasında öngörülemez. Bu belirsizlik mülakatların ve hatta bütün personel seçimi tekniklerinin kestirimci geçerliliğini kısıtlar.

Mülakatlar ayrıca bir psikolojik ön yargılar mayın tarlasıdır. Son yıllarda insanlar mülakatçıların genellikle üstünde düşünmeden, kültürel açıdan kendilerine benzer olan ya da cinsiyet, ırk ve eğitim geçmişi gibi ortak noktalarının olduğu adayları kayırma eğilimi sergilediklerinin fazlasıyla farkındalar. Ön yargılardan kaynaklanan riskleri fark eden pek çok şirket, bu riskleri, işe alımdan sorumlu profesyonellere ve diğer çalışanlara özel eğitim vererek gidermeye çalışıyor. Onlarca yıldır bilinen başka ön yargılar da mevcut. Örneğin adayların değerlendirilmesinde, çok az önem taşıması ya da hiç önemli olmaması gereken pozisyonlarda bile, dış görüntü büyük rol oynuyor. Bu tür ön yargılar bütün ya da çoğu işe alım uzmanları tarafından paylaşılıyor ve belli bir adaya uygulandıklarında, ortak bir hata –adayın değerlendirilmesinde olumlu ya da olumsuz bir ön yargı– üretme olasılıkları çok büyük.

İşe alımda gürültünün de olduğunu duymak sizi şaşırtmayacaktır. Farklı mülakat uzmanları aynı adaya farklı şekilde tepki verir ve farklı sonuçlara varırlar. İki mülakat uzmanının aynı adayla görüşmelerine verdikleri puanların arasındaki korelasyon ölçümleri 0,37 ve 0,44 arasında değişir (YU = %62-65). Bunun bir

nedeni adayın farklı mülakat uzmanlarının karşısında aynı şekilde davranmama olasılığıdır. Ancak, birden fazla mülakat uzmanının aynı aday davranışına maruz kaldığı panel mülakatlarında bile, puanlamalar arasındaki korelasyon kusursuz olmaya çok uzaktır. Bir meta analizi 0,74'lük bir korelasyon (YU = %76) öngörüyor. Bu, sizin ve bir başka mülakat uzmanının *aynı* panel mülakatında *aynı* iki adayı gördükten sonra, hangi adayın daha iyi olduğu konusunda, her dört seferden birinde anlaşmazlık yaşayacağınız anlamına geliyor.

Bu değişkenlik büyük ölçüde örüntü gürültüsünün, mülakatçıların belli bir adaya verdikleri kendilerine has tepkiler arasındaki farkın ürünüdür. Çoğu örgüt bu çeşitliliği bekler ve bu nedenle birden fazla mülakat uzmanının aynı adayla, sonuçların bir şekilde bir araya toplanacağı görüşmeler yapmasını ister. (Tipik olarak bir araya toplanan görüş bir tür uzlaşmaya varılması gereken bir tartışma aracılığıyla şekillenir. Daha önce değindiğimiz gibi, bu da kendi sorunlarını yaratan bir prosedürdür.)

Daha şaşırtıcı bir bulgu, mülakatlarda bol miktarda durum gürültüsünün olmasıdır. Örneğin, işe alma tavsiyelerinin mülakatın adayı rahatlatmak için dostça sohbet edilen iki üç dakikalık gayriresmî ilişki kurma safhasıyla bağlantılı olduğunu gösteren çok güçlü kanıtlar bulunmaktadır. Öyle görünüyor ki ilk izlenim önemli, hem de çok.

Belki de ilk izlenimlere dayanarak bir yargıya varmanın sorun olmadığını düşünüyorsunuz. En azından ilk izlenimden öğrendiklerimizin bir kısmı anlamlıdır. Yeni bir tanıdıkla etkileşimimizin ilk saniyelerinden bir şeyler öğrendiğimizi hepimiz biliriz. Bunun becerikli mülakat uzmanları için özellikle doğru olması son derece mantıklı görünüyor. Ancak bir mülakatın ilk saniyeleri ilk izlenimlerle ilişkilendireceğiniz türde yüzeysel özellikleri yansıtır: İlk algılar çoğu zaman adayın dışa dönüklüğüne ve sözel becerilerine dayanır. El sıkışma şekli bile işe alma tavsiyesi olarak önemli bir

kestiricidir! Candan el sıkışmaları hepimiz seviyor olabiliriz ama candan bir el sıkışmayı kilit bir kriter olarak görmeyi çok az işe alım uzmanı bilinçli olarak tercih edecektir.

Mülakat Uzmanlarının Psikolojisi

İlk izlenimler, neden aslında çok daha uzun süren bir mülakatın sonucunda bu kadar belirleyici olmaktadır? Bunun bir nedeni geleneksel bir mülakatta mülakat uzmanlarının görüşmeyi uygun gördükleri yöne çevirme özgürlüğüne sahip olmasıdır. İlk izlenimlerini teyit edecek sorular sormaları mümkündür. Örneğin bir aday çekingen ve mesafeli göründüğünde mülakatı yapan kişi adayın geçmiş takım çalışması tecrübeleri hakkında zor sorular sormak isteyebilir ama belki de neşeli ve sokulgan görünen bir başka adaya aynı soruları sormayı ihmal edebilir. Bu iki aday hakkında toplanan kanıtlar aynı olmayacaktır. Öz geçmişler ya da test puanlarından olumlu ya da olumsuz bir ilk izlenim oluşturan mülakat uzmanlarının davranışlarını izleyen bir çalışma, ilk izlenimlerin mülakatın ilerleme şekli üzerinde çok etkili olduğu sonucunu ortaya koydu. Olumlu ilk izlenime sahip mülakat uzmanları, örneğin, daha az soru soruyor ve şirketi adaya "satma" eğiliminde oluyor.

İlk izlenimlerin gücü mülakatların tek sorunlu yönü değil. Bir diğer sorun, mülakatı yapan kişi olarak, karşımızda oturan adayın *anlamlı gelmesini* istememiz (Bölüm 13'te bahsettiğimiz aşırı tutarlılık arama ve bulma eğilimimizin bir tezahürü). Çarpıcı bir deneyde araştırmacılar öğrencilerden mülakat uzmanı ve aday rolünü oynamalarını istediler ve her iki tarafa mülakatın kapalı uçlu, evet-hayır sorularından oluşması gerektiğini söylediler. Sonra adaylardan bazılarından soruları *rastgele* yanıtlamalarını istediler. (Evet mi yoksa hayır mı diyeceklerini soruların ilk harfi belirliyordu.) Araştırmacıların istihzayla yazdığı gibi, "Adayların

bazıları başlangıçta rastgele mülakatın dağılacağından ve ortaya bir saçmalık çıkacağından endişeliydi. Böyle bir sorun yaşanmadı ve mülakatlar ilerledi." Doğru okudunuz: *Tek bir mülakatçı bile* adayların rastgele cevap verdiğini fark etmedi. Daha kötüsü, "birlikte geçirdikleri süreye dayanarak kişi hakkında çok çıkarımda bulunup bulunamayacaklarını" tahmin etmeleri istendiğinde bu "rastgele" duruma maruz kalan mülakatçılar, olumlu cevap vermeye hakiki cevaplar veren adaylarla görüşen diğer mülakatçılar kadar yatkındılar. Tutarlılık yaratma becerimiz bu kadar büyüktür. Rastgele veride sıklıkla hayalî bir örüntü bulabildiğimiz ya da bir bulutun hatlarında bir şekil hayal edebildiğimiz gibi, tamamen anlamsız cevaplarda mantık bulmaya da muktediriz.

Daha az çarpıcı bir örnek olarak, şu vakayı düşünün. Bu kitabın yazarlarından birinin, daha önceki pozisyonunda orta ölçekli bir şirkette CFO olan bir adayla mülakat yapması gerekti. Adayın o pozisyondan sadece birkaç ay içinde ayrılmış olduğunu fark ederek ona nedenini sordu. Aday ayrılma sebebinin "CEO ile stratejik bir anlaşmazlık" olduğunu açıkladı. Yazarın bir meslektaşı da aynı adayla mülakat yaptı, aynı soruyu sordu ve aynı cevabı aldı. Ancak bunu izleyen bir fikir alışverişi görüşmesinde, iki mülakatçı çok farklı görüşler belirttiler. Adaya dair olumlu bir izlenim benimsemiş olan mülakatçı adayın bir önceki şirketinden ayrılma kararını bir dürüstlük ve cesaret işareti olarak görüyordu. Olumsuz bir ilk izlenim edinen diğeri ise, aynı olguyu esnek olmama hatta belki de toyluk belirtisi olarak yorumlamıştı. Bu hikâye, bir adayla ilgili yargımızın olgulara dayandığına inanmayı ne kadar istesek de, olguları yorumlama biçimimizin daha önceki tavırlarla renklendiğini gösteriyor.

Geleneksel mülakatların sınırları onlardan anlamlı sonuçlar çıkarma becerimize ciddi bir şüphe düşürür. Bununla birlikte, bir mülakatta biçimlenen izlenimler canlıdır ve mülakatçı genellikle izlenimlerinden emindir. Bir mülakatta ulaşılan sonuçları adayla

ilgili başka işaretlerle birleştirirken, mülakata çok fazla, test puanları gibi, daha kestirimci olabilecek diğer verilere çok az ağırlık vermeye meylederiz.

Bir hikâye bu gözlemin gözünüzde canlanmasına yardımcı olabilir. Üniversitede bir kadro için mülakat yapan hocaların öğretme becerilerinin kurumun standartlarına uygun olduğundan emin olmak için, genellikle meslektaşlarından oluşan bir heyetin karşısında ders anlatmaları istenir. Elbette sıradan bir sınıf karşısında ders anlatmaktan daha riskli bir durumdur. Aramızdan biri bir keresinde bir adayın bu egzersizde, muhtemelen durumun stresi nedeniyle kötü bir izlenim bıraktığına şahit oldu. Adayın öz geçmişinde öne çıkan öğretme değerlendirmelerinden ve öğretmenlik becerisi nedeniyle aldığı ödüllerden bahsediliyordu. Buna rağmen, adayın, fazlasıyla yapay bir durumda yaşanan başarısızlığının ürettiği canlı izlenim, nihai kararda geçmişteki mükemmel öğretmenlik performansıyla ilgili soyut verilerden çok daha etkili oldu.

Son bir husus: Mülakatlar adaylarla ilgili tek bilgi kaynağı olmadığında –örneğin, ayrıca testler, referanslar ve diğer girdiler olduğunda–, bu çeşitli girdiler bir araya gelerek toplam bir yargı oluşturabilir. Bu durumun akla getirdiği soruyu artık tanıyorsunuz: Girdiler yargı kullanılarak (klinik toplamayla) mı, yoksa formül kullanılarak (mekanik toplamayla) mı birleştirilmeli? Bölüm 9'da gördüğümüz gibi, mekanik yaklaşım hem genel anlamda hem de iş performansı tahmini örneğinde daha üstündür. Ne yazık ki anketler, İK profesyonellerinin ezici çoğunluğunun klinik toplamayı yeğlediklerini gösteriyor. Bu uygulama zaten gürültülü olan bir sürece bir gürültü kaynağı daha ekliyor.

Personel Seçimini Yapılandırmayla İyileştirme

Geleneksel mülakatlar ve yargıya dayalı işe alma kararları kestirimci geçerliliği sınırlandırdıysa, bu konuda ne yapabiliriz? Neyse ki

araştırmalar personel seçiminin nasıl iyileştirilebileceği konusunda birtakım tavsiyeler de üretti ve bazı firmalar bu tavsiyelere uyuyor.

Personel seçimi uygulamalarını geliştiren ve sonuçlarını bildiren şirketlerin bir örneği Google. Google'ın eski İnsan Operasyonları'ndan Sorumlu Kıdemli Başkan Yardımcısı Laszlo Bock hikâyeyi *Google Sırları* adlı kitabında anlatıyor. En yüksek kalibrede yetenekleri işe almaya odaklanmalarına ve doğru kişileri bulmaya azımsanmayacak bir kaynak ayırmalarına rağmen, Google sıkıntı yaşıyordu. İşe alma mülakatlarının kestirimci geçerliliğini ölçmek için yapılan bir denetim "sıfır ilişki... tamamen gelişigüzel bir kargaşa", saptadı. Google'ın bu duruma çare bulmak için benimsediği değişiklikler onlarca yıllık araştırmalardan doğan prensipleri yansıtıyordu. Ayrıca karar hijyeni stratejilerini de temsil ediyorlardı.

Bu stratejilerden birine artık aşina olmalısınız: bir araya toplama. Bu bağlamda kullanılması sürpriz değil. Neredeyse bütün şirketler birden fazla mülakat uzmanının aynı adayla ilgili yargılarını bir araya toplar. Altta kalmamak için, Google zaman zaman bazı adayları yirmi beş mülakattan geçiriyordu! Bock'un incelemesinin sonuçlarından biri bu sayıyı dörde indirmek oldu; çünkü ek mülakatların ilk dört mülakatta edinilen kestirimci geçerliliğe neredeyse hiçbir katkısı olmadığını görmüştü. Yine de, Google bu geçerlilik seviyesini sağlama almak için, bütün şirketlerde uygulanmayan bir kuralı katı bir şekilde uyguluyordu: Şirket, mülakat uzmanlarının adayı, birbirleriyle iletişim kurmadan *önce*, ayrı ayrı puanladıklarından emin oluyordu. Bir kez daha, bir araya toplama işe yarar ama sadece yargılar bağımsız olduğunda.

Google ayrıca henüz detaylı olarak anlatmadığımız bir diğer karar hijyeni stratejisini de benimsedi: *karmaşık yargıları yapılandırma. Yapılandırma* terimi pek çok anlama gelebilir. Bizim burada kullandığımız hâliyle, yapılandırılmış bir karmaşık yargıyı üç prensiple tanımlıyor: parçalama, bağımsızlık ve geciktirilmiş bütüncül yargı.

İlk prensip olan *parçalama* kararı ögelerine ya da *aracı değerlendirmelere* ayırıyor. Bu adım, alt yargıların bir kılavuzda belirlenmesiyle aynı amacı taşıyor: Karar vericileri önemli işaretlere odaklıyor. Parçalama hangi verilere ihtiyaç olduğunu belirleyen bir yol haritası görevi üstleniyor. Ve alakasız bilgiyi filtre ediyor.

Google örneğinde parçalamada dört aracı değerlendirme var: genel bilişsel beceri, liderlik, kültürel uyum ("googleyness" deniyor) ve rolle ilgili bilgi. (Daha sonra bu değerlendirmelerin bazıları daha küçük ögelere ayrılıyor.) Bir adayın dış görüntüsünün, hoşsohbet olmasının ve heyecan verici hobilerinin ve işe alım yapan kişinin yapılandırılmamış bir mülakatta dikkat edebileceği olumlu ya da olumsuz diğer yönlerinin listede yer almadığına dikkatinizi çekmek isteriz.

İşe alım için böyle bir yapı yaratmak mantıklı görünebilir. Hatta giriş seviyesinde bir muhasebeci ya da bir yönetici asistanı alıyorsanız, ihtiyaç duyulan becerileri belirleyen standart iş tanımları vardır. Öte yandan profesyonel işe alım uzmanlarının bildiği gibi, kilit değerlendirmeleri tanımlamak sıra dışı ya da kıdemli pozisyonlarda zorlaşır ve bu tanım adımı genellikle atlanır. Saygın ve önemli bir yetenek avcısı istenen becerilerin yeterince belirli şekilde tanımlanmasının zor ve sık sık pas geçilen bir görev olduğunu belirtiyor. Karar alıcıların "sorunun tanımına yatırım yapması"nın, önceden, herhangi bir adayla tanışmadan önce, net ve detaylı bir iş tanımı üzerinde anlaşmaya gerekli zamanı ayırmasının öneminin altını çiziyor. Buradaki zorluk pek çok mülakatçının uzlaşma ve ödünle üretilmiş şişirilmiş iş tanımlarını kullanması. Tanımlar ideal bir adayın sahip olacağı bütün özelliklerin muğlak bir dilek listesi gibi ve özellikleri kalibre etmek ya da aralarında değiş tokuş yapmak için bir yol önermiyorlar.

Yapılandırılmış yargının ikinci prensibi olan *bağımsızlık* her değerlendirmedeki bilginin bağımsız olarak toplanmasını gerektiriyor. İş tanımının ögelerini listelemek yeterli değil: Geleneksel

mülakatlar yapan çoğu işe alım uzmanı adayda aradıkları dört ya da beş özelliği de bilirler. Sorun şu ki mülakat sırasında bu ögeleri ayrı ayrı değerlendirmezler. Her değerlendirme diğerlerini etkiler ve bu da her bir değerlendirmeyi çok gürültülü kılar.

Bu sorunu aşmak için, Google olguya dayalı ve birbirinden bağımsız değerlendirmeler yapmanın yollarını geliştirdi. Belki de en gözle görülür hamleleri *yapılandırılmış davranışsal mülakatlar* kavramını devreye sokmaları oldu. Bu tür mülakatlarda mülakatçının görevi bir adayı genel olarak sevip sevmediklerine karar vermek değil, değerlendirme yapısındaki her başlık için veri toplamak ve her başlık için adaya bir not vermek. Bunu yapmak için mülakat uzmanlarından adayın geçmiş durumlardaki davranışları hakkında önceden belirlenmiş sorular sormaları isteniyor. Ayrıca cevapları kaydetmeleri ve onları, önceden belirlenmiş bir puanlama ölçeğine göre, birleşik bir yönerge kullanarak puanlamaları gerekiyor. Yönerge her soru için ortalama, iyi ve harika cevaplarının neye benzediği konusunda örnekler veriyor. Bu ortak ölçek (bir önceki bölümde bahsettiğimiz davranışsal olarak çapalı puanlama ölçeklerinin bir örneği) yargılardaki gürültünün azaltılmasına yardımcı oluyor.

Bu yaklaşım size geleneksel, bol sohbetli mülakatlardan farklı göründüyse, gerçekten öyle. Aslında bir iş görüşmesinden çok, bir sınav ya da sorgu hissi vermesi daha mümkün ve her iki tarafın da yapılandırılmış mülakatlardan hoşlanmadığına (ya da en azından yapılandırılmamış olanları yeğlediklerine) dair kanıtlar hiç de az değil. Bir mülakatın yapılandırılmış olarak kabul edilmesi için neleri içermesi gerektiği konusunda tartışmalar devam ediyor. Yine de, mülakatları konu alan literatürden çıkan en tutarlı bulgulardan biri, yapılandırılmış mülakatların gelecekteki iş performansı konusunda, geleneksel ve yapılandırılmamış mülakatlardan çok daha kestirimci olduğu. İş performansıyla korelasyonlar 0,44 ve 0,57 arasında değişiyor. YU ölçümümüzü kullanırsak, yapılandı-

rılmış bir mülakatta daha iyi adayı seçme olasılığınız %65 ile %69 arasında; bu yapılandırılmamış bir mülakatın size vereceği %56 ve %61 şanstan çok daha yüksek.

Google önemsediği bazı boyutlarda girdi olarak başka verileri de kullanıyor. İşle ilgili bilgiyi test etmek için, kısmen *iş örneği testleri*nden faydalanıyor: bir programlama işi için adaydan bir kod yazmasını istemek gibi. Araştırmalar iş örneği testlerinin işteki performans konusunda en iyi kestiricilerden biri olduğunu gösteriyor. Google ayrıca adayın kendisinin seçtiği kişilerden sağlanan değil, adayın daha önce yolunun kesiştiği Google çalışanlarından alınan "arka kapı referanslarını" da değerlendiriyor.

Yapılandırılmış yargının üçüncü prensibi, *geciktirilmiş bütüncül yargı*, basit bir reçeteyle tarif edilebilir. Sezgiyi tamamen dışlamayın, ama geciktirin. Google'da son işe alma tavsiyesi eşit görüş ağırlığı olan bir komite tarafından gerçekleştirilir. Komite, adayların her mülakattaki değerlendirmede aldığı bütün puanları ve bu değerlendirmeleri destekleyen diğer ilgili bilgileri içeren eksiksiz bir dosyayı gözden geçirir. Daha sonra bu bilgilere dayanarak bir teklif yapıp yapmamaya karar verir.

Bu şirketin herkesçe bilinen veriye dayalı kültürüne rağmen ve verilerin mekanik kombinasyonunun klinik kombinasyondan daha iyi sonuç verdiği yönündeki bütün kanıtlara rağmen, nihai işe alma kararı mekanik *değildir.* Komitenin bütün kanıtları hesaba kattığı ve bütüncül olarak tarttığı ve "Bu kişi Google'da başarılı olur mu?" sorusunun tartışıldığı bir yargı olma özelliğini korur. Karara sadece hesapla ulaşılmaz.

Bir sonraki bölümde, nihai karar almaya bu yaklaşımı neden mantıklı bulduğumuzu açıklayacağız. Ama unutmayın ki Google'un nihai işe alma kararları, mekanik olmamakla birlikte, dört mülakatçının verdiği ortalama puandan yola çıkar. Ayrıca altta yatan kanıtlardan da yararlanır. Başka bir deyişle, Google karar alma sürecinde yargı ve sezgiye ancak bütün kanıtlar toplanıp

analiz edildikten sonra izin verir. Dolayısıyla, her mülakatçının (ve işe alma komitesi üyelerinin) hızlı ve sezgisel izlenim oluşturma ve yargıya varmak için acele etme eğilimi kontrol altında tutulur.

Üç prensip –tekrar ediyoruz, parçalama, her boyutu bağımsız değerlendirme ve geciktirilmiş bütüncül yargı– seçme süreçlerini iyileştirmeye çalışan bütün örgütler için illa bir şablon oluşturacak diye bir şey yok. Ancak prensipler örgütsel psikologların yıllar içinde şekillendirdiği tavsiyelerle büyük ölçüde tutarlı görünüyor. Hatta prensipler içimizden birinin (Kahneman) ta 1956'da İsrail ordusunda uyguladığı ve *Hızlı ve Yavaş Düşünme* kitabında anlattığı seçme metoduyla birtakım benzerlikler taşıyor. O süreç, tıpkı Google örneği gibi, bir değerlendirme yapısını (değerlendirilmesi gereken kişilik ve beceri boyutlarını listesini) şekillendirmişti. Mülakatçıların her boyutla alakalı objektif kanıt edinmesini ve bir sonrakine geçmeden önce, o boyutu puanlamasını gerektiriyordu. Ve göreve alım yapan kişilerin nihai bir karara varmak için yargı ve sezgi kullanmasına izin veriyordu, ancak bunun için önce yapılandırılmış değerlendirmenin tamamlanması şarttı.

İşe alımda yapılandırılmış yargı süreçlerinin (yapılandırılmış mülakatlar dâhil) üstünlüğü konusunda sarsıcı kanıtlar var. Bu süreçleri benimsemek isteyen üst düzey yöneticilere rehberlik edecek pratik tavsiyeler mevcut. Google örneğinin gösterdiği ve diğer araştırmacıların ifade ettiği gibi, yapılandırılmış yargı yöntemleri aynı zamanda daha az maliyetliler çünkü yüz yüze görüşmeye ayrılan zaman kadar maliyetli şey çok az.

Öte yandan, çoğu üst düzey yönetici mülakata dayalı, gayriresmî yöntemlerin yeri doldurulamaz değerine inancını koruyor. Dikkat çekici bir şekilde, bu, potansiyel bir işverene gerçek azimlerini sadece yüz yüze mülakatla gösterebileceklerine inanan adaylar için de geçerli. Araştırmacılar buna "bir yanılsamanın ısrarı" adını

veriyor. Kesin olan bir şey var: İşe alımcılar ve adaylar işe alma yargılarında gürültünün önemini ciddi derecede hafife alıyorlar.

İşe Alımda Yapıdan Bahsederken

"Geleneksel, gayriresmî mülakatlarda, sıklıkla adayı anlamak ve kişinin pozisyona uyup uymadığını bilmek için karşı konulmaz, sezgisel bir hissimiz olur. Bu hisse güvenmemeyi öğrenmeliyiz."

"Geleneksel mülakatlar sadece ön yargılar nedeniyle değil, gürültü nedeniyle de tehlikelidir."

"Mülakatlarımıza ve daha geniş anlamda, seçme süreçlerimize yapı katmalıyız. İşe adaylarda aradıklarımızı daha net ve belirgin bir şekilde tanımlamakla başlayalım ve adayları bu boyutların her birinde bağımsız olarak değerlendirdiğimizden emin olalım."

BÖLÜM 25

Aracı Değerlendirmeler Protokolü

Bir süre önce, ikimiz (Kahneman ve Sibony) dostumuz Dan Lovallo'yla birlikte, örgütlerde bir karar alma yöntemi tanımladık. Gürültüyü hafifletmeyi öncelikli amaç edinen bu yönteme, *aracı değerlendirmeler protokolü* adını verdik. Yöntem daha önceki bölümlerde değindiğimiz karar hijyeni stratejilerinin büyük kısmını kapsıyor. Bu protokol geniş bir alanda ve bir planın ya da seçeneğin değerlendirilmesinin çoklu boyutları göz önünde bulundurmayı ve tartmayı gerektirdiği durumlarda uygulanabilir. Çeşitli şirketler, üniversiteler ve hükûmet bürolarının aralarında olduğu her tür örgüt tarafından, çok çeşitli şekillerde kullanılıp uyarlanabilir.

Burada protokolü birkaç gerçek vakanın bir bileşimi olan stilize edilmiş bir örnekle, Mapco adını verdiğimiz kurgusal bir şirket üstünden anlatacağız. Mapco'nun çok önemli ve dönüştürücü bir şirket alımı fırsatını değerlendirirken izlediği adımları inceleyecek ve bu adımların bir şirketin benzer bir durumda attığı alışılmış adımlardan farklarını göreceğiz. Sizin de göreceğiniz gibi, aradaki farklar önemli ancak dikkatsiz bir gözlemcinin fark etmeyebileceği kadar incelikli.

İlk Toplantı: Yaklaşım Konusunda Anlaşmak

Rakip bir firma olan Roadco'yu satın alma fikri Mapco'ya cazip geliyordu ve yeterince olgunlaştığı için şirketin liderleri konuyu tartışmak için bir yönetim toplantısı yapmayı düşünüyorlardı. Mapco'nun CEO'su Joan Morrison, olası bir satın almanın ve kurulun bu konudaki müzakerelerini iyileştirmek için nelerin yapılması gerektiğinin bir ön görüşmesi için, yönetim kurulunun strateji komitesini topladı. Toplantının başında Joan komiteyi bir teklifle şaşırttı:

"Roadco'nun alımınına karar vereceğimiz yönetim kurulu toplantısında yeni bir prosedür denememizi önermek istiyorum. Bu yeni prosedürün itici bir adı var: Aracı değerlendirme protokolü ama fikir hayli basit. Stratejik bir seçeneğin değerlendirilmesiyle, bir iş adayının değerlendirmesi arasındaki benzerlikten ilham alıyor.

Yapılandırılmış mülakatların yapılandırılmamış mülakatlardan daha iyi sonuçlar ürettiğini gösteren araştırmaya ve daha geniş anlamda, yapılandırmanın işe alım kararını iyileştirdiği fikrine aşina olduğunuzu tahmin ediyorum. İK departmanımızın işe alım kararlarında bu prensipleri benimsediğini biliyorsunuz. Çok sayıda araştırma mülakatlarda yapının daha yüksek isabetliliğe yol açtığını, eskiden bizim de uyguladığımız yapılandırılmamış mülakatların bu isabetliliğin yakınından bile geçmediğini gösteriyor.

Adayların değerlendirilmesiyle, büyük kararlarda seçeneklerin değerlendirilmesi arasında bariz bir benzerlik görüyorum: *Seçenekler adaylar gibiler.* Ve bu benzerlik beni, adayları değerlendirmede işe yarayan bu yöntemi, önümüzdeki göreve, yani stratejik seçenekleri değerlendirmeye uyarlama fikrine yönlendiriyor."

Bu benzetme, başlangıçta komite üyelerinin kafasını karıştırdı. İşe alma sürecinin sayısız benzer karar üreten iyi işleyen bir mekanizma olduğunu ve bunu ciddi zaman baskısı altında yapmadığını savundular. Öte yandan, stratejik bir karar bol miktarda o amaca özgü çalışma gerektirirdi ve çabuk alınması gerekirdi.

Bazı komite üyeleri Joan'a kararı geciktirecek her teklife karşı olacaklarını açıkça belirttiler. Ayrıca, Mapco'nun araştırma personelinin durum değerlendirme gerekliliklerine ekleme yapmak konusunda da endişeliydiler.

Joan bu itirazları direkt olarak cevapladı. Çalışma arkadaşlarına yapılandırılmış sürecin kararı geciktirmeyeceği konusunda güvence verdi. "Bütün mesele, anlaşmayı tartışacağımız yönetim kurulu toplantısı için gündemi hazırlamak", diye açıkladı. "Anlaşmanın farklı yönlerinin değerlendirme listesine önceden karar vermeliyiz; tıpkı bir mülakat uzmanının işe adayın sahip olması gereken özelliklerin kontrol listesi olma görevini üstlenen bir iş tanımıyla başlaması gibi. Yönetim kurulunun bu değerlendirmeleri ayrı ayrı, tek tek tartışmasını sağlayacağız; tıpkı yapılandırılmış mülakatlarda görüşmeyi yapanların adayı farklı boyutlar için sırayla değerlendirmesi gibi. Anlaşmayı kabul etme ya da reddetme tartışmasına, ancak bunu yaptıktan sonra döneceğiz. Bu süreç yönetim kurulunun kolektif bilgeliğinden faydalanmanın çok daha etkili bir yolu olacak.

Bu yaklaşım üzerinde anlaşırsak, elbette, bilginin nasıl sunulması gerektiği ve anlaşma takımının toplantıyı hazırlarken nasıl çalışması gerektiği konusunda etkileri olacak. İşte bu yüzden fikirlerinizi şimdi almak istedim."

Komite üyelerinden hâlâ şüpheci olan biri, Joan'a yapının işe alma kararlarının kalitesine ne gibi faydalar sağladığını ve bu faydaların stratejik bir karara aktarılabileceğine neden inandığını sordu. Joan ona düşüncenin mantığını adım adım izah etti. Aracı değerlendirmeler protokolünü kullanmanın, değerlendirmenin boyutlarını birbirinden bağımsız tutarak bilginin değerini en üst düzeye çıkardığını anlattı. "Normalde yönetim kurulunda yaptığımız tartışmalar, yapılandırılmamış mülakatlara çok benziyor." dedi. "Bir karar varma amacının sürekli farkındayızdır ve bütün bilgiyi o amacın ışığında işleriz. İşe bir kapanış arayarak başlar

ve buna olabildiğince çabuk ulaşırız. Yapılandırılmamış bir mülakattaki bir işe alımcı gibi, bütün müzakereyi ilk izlenimlerimizi teyit etmek için kullanma riskiyle karşı karşıya kalırız."

"Yapılandırılmış bir yaklaşım kullanmak bizi, bir karara varma amacını bütün değerlendirmeleri tamamlayana kadar ertelemeye mecbur bırakacak. Ayrı değerlendirmeleri ara hedefler olarak göreceğiz. Bu şekilde, erişebileceğimiz bütün bilgiyi göz önüne alacak ve anlaşmanın bir yönündeki çıkarımımızın alakasız ve farklı bir yönündeki yorumumuzu değiştirmediğinden emin olacağız."

Komite yaklaşımı denemeyi kabul etti. Fakat söz konusu aracı değerlendirmeler nelerdi? Joan'ın aklında önceden belirlenmiş bir kontrol listesi var mıydı? Joan, "Hayır." dedi. "Protokolü rutin bir karara uygulasaydık, olabilirdi ama bu örnekte, aracı değerlendirmeleri kendimiz tanımlamalıyız. Bu kritik önem taşıyor: Değerlendirilmesi gereken şirket alımının başlıca yönlerine karar vermek bize düşüyor." Strateji komitesi ertesi gün bunu yapmak için toplanmaya karar verdi.

İkinci Toplantı: Aracı Değerlendirmeleri Tanımlamak

Joan, "Yapacağımız ilk iş," diye açıkladı, "anlaşmayla ilgili bağımsız değerlendirmelerin kapsamlı bir listesini çıkarmak. Bunlar Jeff Schneider'in araştırma ekibi tarafından değerlendirilecek. Bizim bugünkü görevimiz değerlendirmelerin listesini oluşturmak. Liste aklınıza gelen her tür alakalı olgunun yer alacağı ve değerlendirmelerden en az birini etkileyeceği şekilde kapsamlı olmalı. Ve 'bağımsız'dan kastım, fazlalıkları en aza indirmek için, alakalı bir olgunun tercihen değerlendirmelerden sadece birini etkilemesi."

Grup işe koyuldu ve alakalı görünen olgu ve verilerden uzun bir liste üretti. Daha sonra o listeyi bir değerlendirmeler listesi şeklinde düzenlediler. Katılımcıların çok geçmeden fark ettiği gibi, işin en zor kısmı listenin kısa ve kapsamlı olmasını ve üst

üste binmeyen değerlendirmelerden oluşmasını sağlamaktı. Ama altından kalkılabilecek bir görevdi. Hatta grubun yedi değerlendirmeden oluşan nihai listesi yüzeysel olarak yönetim kurulunun bir şirket alımı teklifini sunan normal bir rapordan bekleyeceği içindekiler kısmına benziyordu. Beklenen finansal modellemeye ek olarak, liste örneğin hedefin yönetim ekibinin kalitesinin değerlendirmesini ve beklenen sinerjilerin yakalanma olasılığının değerlendirmesini de içeriyordu.

Strateji komitesi üyelerinden bazıları toplantının Roadco hakkında yeni içgörüler üretmemesi karşısında hayal kırıklığına uğradılar. Ama Joan amacın bu olmadığını anlattı. Yakın hedef, şirket alımını incelemekle sorumlu anlaşma ekibini bilgilendirmekti. Her değerlendirme, anlaşma ekibinin raporunda ayrı bir bölümün konusu olacak ve yönetim kurulu tarafından ayrı olarak tartışılacaktı.

Anlaşma takımının görevi, Joan'ın bakış açısıyla, yönetim kuruluna anlaşmanın bütünü hakkındaki düşüncesini sunmak değildi. En azından, şimdilik. Takımın görevi aracı değerlendirmelerin her biri hakkında, objektif ve bağımsız bir değerlendirme sağlamaktı. Joan son olarak, anlaşma ekibinin raporundaki her bölümün basit bir soruyu cevaplayan bir puanlamayla bitmesi gerektiğini açıkladı: "Bu konuya nihai kararda vermemiz gereken ağırlığı bir kenara bırakarak, bu değerlendirmeyle ilgili kanıtlar anlaşmanın ne kadar lehine ya da aleyhine bir savunu getiriyor?"

Anlaşma Ekibi

Anlaşmanın değerlendirilmesinden sorumlu ekibin lideri Jeff Schneider o öğleden sonra çalışmayı organize etmek için ekibini bir araya topladı. Ekibin her zamanki çalışma şeklinden farklılıklar çok fazla değildi ama Schneider önemlerini özellikle vurguladı.

Öncelikle, ekibin analizcileri analizlerini olabildiğince objektif tutmaya gayret etmeliydiler. Analizler olgulara dayanmalıydı –bunda yeni bir şey yoktu– ama aynı zamanda mümkün olan her adımda bir *dış bakış* kullanılmalıydı. Ekip üyeleri "dış bakış"la neyi kastettiğinden emin olamayınca, Jeff onlara Joan'ın belirttiği aracı değerlendirmelerden ikisini kullanarak iki örnek verdi. Anlaşmanın düzenleyici onay alma olasılığını değerlendirmek için, işe *baz oranı,* yani onaylanan karşılaştırılabilir işlemlerin oranını bulmakla başlamaları gerektiğini belirtti. Bu görev daha sonra ilgili bir *referans sınıfı,* yani yeterince karşılaştırılabilir anlaşmalardan oluşan bir grup tanımlamalarını gerektirecekti.

Jeff daha sonra hedef şirketin ürün geliştirme departmanının teknolojik becerilerinin nasıl değerlendirileceğini anlattı. Bu da Joan'ın listelediği bir diğer önemli değerlendirmeydi. "Şirketin yakın zamandaki başarılarını olguya dayalı olarak tanımlamak ve 'iyi' ya da 'harika' olarak nitelemek yeterli değil. Sizden beklediğim şuna benzer bir şey: "'Bu ürün geliştirme departmanı, yakın zamandaki ürün lansmanı kayıtlarıyla ölçüldüğü üzere, emsal grubunun ikinci yüzde yirmilik diliminde yer alıyor.'" Jeff toplamda, amacın olabildiğince karşılaştırıcı değerlendirmeler yapmak olduğunu açıkladı, çünkü göreceli yargılar mutlak yargılardan daha iyiydi.

Jeff'in bir talebi daha vardı. Joan'ın talimatlarına göre, bir değerlendirmenin bir diğerini etkileme riskini azaltmak için değerlendirmeler birbirlerinden olabildiğince bağımsız olmalıydı. Buna göre, farklı analizcileri farklı değerlendirmelerde görevlendirdi ve bağımsız çalışmalarını istedi.

Analizcilerden bazıları şaşkınlıklarını ifade ettiler. "Takım çalışması daha iyi değil mi?" diye sordular. "İletişim kurmamızı istemiyorsanız, bir ekip kurmanın ne anlamı var?"

Jeff bağımsızlık ihtiyacını açıklaması gerektiğinin farkındaydı. "İşe alımda halo etkisini muhtemelen biliyorsunuzdur." dedi. "Bir adayın genel izleniminin, adayın belirli bir boyuttaki becerilerini

değerlendirmenizi etkilediği zaman olan şey. Kaçınmaya çalıştığımız şey bu." Analizcilerin bazıları bu etkinin ciddi bir sorun olduğunu düşünmüyor gibi göründüğü için, Jeff başka bir benzetme kullandı: "Bir suçun dört tanığı varsa, ifade vermeden önce konuşmalarını ister misiniz? Elbette hayır! Bir tanığın diğerlerini etkilemesini istemezsiniz." Analizciler bu karşılaştırmayı pek hoş bulmadılar ama Jeff gerekli mesajın iletildiğini düşünüyordu.

Gelin görün ki Jeff'in tamamen bağımsız değerlendirme hedefine ulaşmak için yeterli analizcisi yoktu. Ekibin tecrübeli üyelerinden Jane iki değerlendirmede görevlendirilmişti. Jeff bu iki değerlendirmeyi, birbirlerinden olabildiğince farklı olacak şekilde seçti ve Jane'e ilk değerlendirmeyi tamamlayıp raporunu hazırlamadan, ikinci değerlendirmeye geçmemesi için talimat verdi. Bir diğer kaygı yönetim ekibinin kalitesinin değerlendirilmesiydi. Jeff analizcilerin yönetim ekibinin asli kalitesiyle ilgili değerlendirmelerini şirketin yakın zamanda aldığı sonuçlardan (ki elbette analiz ekibi bu sonuçları detaylı olarak inceleyecekti) ayırmakta zorlanacaklarından endişeliydi. Jeff bu meseleyi halletmek için, dışarıdan bir İK uzmanından yönetim ekibinin kalitesi konusunda devreye girmesini istedi. Böylece, daha bağımsız bir girdi elde edeceğini umuyordu.

Jeff'in ekibin sıra dışı bulduğu bir talimatı daha vardı. Her bölüm bir değerlendirmeye odaklanmalıydı ve Joan'ın talep ettiği gibi puanlama şeklinde bir sonuca varmalıydı. Ancak, analizciler her bölüme değerlendirmeyle ilgili bütün olgusal bilgiyi mutlaka katmalıydılar. Jeff, "Hiçbir şeyi gizlemeyin", diye tembihledi. "Bölümün genel tonlaması elbette önerilen puanlamayla tutarlı olacak ama eğer ana puanlamayla tutarsız ve hatta çelişkili görünen bilgiler varsa, sakın halının altına süpürmeyin. Sizin işiniz tavsiyenizi pazarlamak değil. Gerçeği temsil etmek. Karmaşık olacaksa da varsın olsun, genellikle öyle olur zaten."

Aynı yaklaşımla, Jeff analizcileri her değerlendirmedeki eminlik seviyeleri konusunda şeffaf olmaya cesaretlendirdi. "Yönetim kurulu elinizde kusursuz bilgi olmadığını biliyor; gerçekten karanlıkta hissettiğiniz zaman bunu onlara söylemeniz faydalı olacaktır. Ve sizi gerçekten duraksatan bir şeye –potansiyel bir iş bozucuya– rastlarsanız, elbette bunu derhâl bildirmelisiniz."

Anlaşma ekibi talimatlara göre ilerledi. Neyse ki önemli iş bozucu hususlara rastlamadılar. Ekip Joan ve yönetim kurulu için yapılan bütün değerlendirmeleri kapsayan bir rapor hazırladı.

Karar Toplantısı

Joan karar toplantısına hazırlanmak için raporu okurken, çok önemli bir şeyi derhâl fark etti: Değerlendirmelerin çoğu anlaşmanın yapılmasını desteklerken, basit, iyimser ve her şeyin yolunda olduğu bir tablo çizmiyorlardı. Puanlamaların bazıları güçlü, diğerleri değildi. Joan bu farklılıkların değerlendirmeleri birbirlerinden bağımsız tutmanın öngörülebilir bir sonucu olduğunu biliyordu. Aşırı tutarlılık kontrol altına alındığı zaman, gerçek durum çoğu yönetim kurulu sunumunun gösterdiği kadar tutarlı olmaz. Joan, "İyi," diye düşündü. "Değerlendirmeler arasındaki bu uyuşmazlıklar soruları doğuracak ve tartışmaları tetikleyecek. Yönetim kurulunda sağlam bir münazara için buna ihtiyacımız var. Sonuçların çeşitliliği elbette kararı kolaylaştırmayacak ama daha iyi kılacağı kesin."

Joan raporu gözden geçirmek ve bir karara varmak için yönetim kurulunu topladı. Anlaşma ekibinin izlediği yaklaşımı açıkladı ve yönetim kurulu üyelerini aynı prensibi uygulamaya davet etti. "Jeff ve ekibi değerlendirmelerin birbirlerinden bağımsız olması için çok çalıştılar," dedi, "ve şimdi bizim görevimiz de onları bağımsız olarak gözden geçirmek. Bu, nihai kararı tartışmaya başlamadan önce, her değerlendirmeyi ayrı ayrı dikkate alacağımız anlamına

geliyor. Her değerlendirmeyi ayrı bir gündem maddesi gibi ele alacağız."

Yönetim kurulu üyeleri bu yapılandırılmış yaklaşımı izlemenin zor olacağını biliyordu. Joan onlardan bütün değerlendirmeler tartışılmadan bütüncül bir görüş oluşturmamalarını istiyordu ama bu insanların çoğu sektörün içindendiler. Roadco hakkında bir görüşleri *vardı.* O görüşten bahsetmemek biraz yapay geliyordu. Bununla birlikte, Joan'ın ne yapmaya çalıştığını anladıkları için, onun kurallarına göre oynamayı ve genel görüşlerini tartışmaktan şimdilik uzak durmayı kabul ettiler.

Yönetim kurulu üyeleri bu uygulamanın son derece değerli olduğunu şaşırarak gördüler. Toplantı sırasında, anlaşmayla ilgili fikri değişenler bile oldu (gerçi bunu kimse bilmeyecekti, çünkü görüşlerini kendilerine sakladılar). Joan'ın toplantıyı yönetme şekli çok büyük rol oynadı: Müzakerenin ve bağımsız görüşlerin ortalamasını almanın avantajlarını birleştiren *tahmin-konuşma-tahmin* yöntemini uyguladı.

Şu şekilde ilerledi: Her değerlendirmede, Jeff anlaşma ekibi adına kilit olguları kısaca özetledi (yönetim kurulu üyeleri daha önce detaylı olarak okumuşlardı). Daha sonra Joan yönetim kurulu üyelerinden telefonlarındaki bir oylama uygulamasını kullanarak değerlendirmeyle ilgili puanlarını vermelerini istedi. Bu puan anlaşma ekibinin önerdiğiyle aynı da olabilirdi, farklı da. Oyların dağılımı, değerlendiricilerin kimlikleri verilmeden hemen ekrana yansıtıldı. "Bu bir oylama değil." diye açıkladı Joan. "Burada sadece her konu için odanın sıcaklığını ölçüyoruz." Joan bir tartışmaya başlamadan önce, her kurul üyesinin bağımsız görüşünü alarak, sosyal etki ve bilgi şelaleleri tehlikesini azaltıyordu.

Bazı değerlendirmelerde, hemen uzlaşmaya varılırken, diğerlerinde süreç zıt görüşleri ortaya serdi. Doğal olarak Joan tartışmaların ikincisine daha çok zaman ayırmasını sağladı. İki taraftaki kurul üyelerinin söz almasını sağlayarak, onları görüşlerini olgu

ve argümanların yanı sıra nüans ve tevazuyla ifade etmeye teşvik etti. Anlaşma konusunda çok güçlü hisleri olan bir yönetim kurulu üyesi kendini kaptırınca ona, "hepimiz aklı başında insanlarız ve görüş ayrılığımız var, bu nedenle bu, aklı başında insanların görüş ayrılığı yaşayabileceği bir konu olmalı." diye hatırlattı.

Bir değerlendirmenin tartışması sona yaklaşırken, Joan kurul üyelerinden puanlarını yeniden oylamalarını istedi. Çoğunlukla ilk turdan daha fazla yakınlaşma oldu. Aynı sıralama –bir ilk tahmin, tartışma ve ikinci bir tahmin– her değerlendirme için tekrar edildi.

Nihayet, anlaşma konusunda bir sonuca ulaşma zamanı geldi. Jeff tartışmayı kolaylaştırmak için değerlendirme listesini, her değerlendirme için kurulun verdiği ortalama puan eşliğinde beyaz tahtada gösterdi. Kurul üyeleri anlaşmanın profiline bakıyorlardı. Nasıl karar vermeliydiler?

Bir kurul üyesinin basit bir önerisi oldu: puanlamaların düz ortalamasını almak. (Belki de Bölüm 9'da bahsedildiği gibi, mekanik toplamanın, bütüncül, klinik yargıya üstünlüğünden haberdardı.) Ancak başka bir üye, buna hemen itiraz etti; ona göre, değerlendirmelerin bazılarına diğerlerinden daha fazla ağırlık verilmeliydi. Üçüncü bir üye de buna itiraz ederek farklı bir değerlendirme hiyerarşisi önerdi.

Joan tartışmanın arasına girdi. "Burada mesele değerlendirme puanlarının basit bir kombinasyonunu hesaplamaktan ibaret değil." dedi. "Sezgiyi geciktirdik ama şimdi kullanma zamanı. Şu anda sizin yargınıza ihtiyacımız var."

Joan mantığını açıklamadı ama dersini zor yoldan almıştı. Özellikle önemli kararlarda insanların ellerini kollarını bağlayan ve yargılarını kullanmalarına izin vermeyen düzenleri reddettiğini biliyordu. Karar alıcıların, bir formülün kullanılacağını bildikleri zaman sistemi nasıl manipüle ettiklerini görmüştü. Puanlamaları arzulanan sonuca ulaşacak şekilde değiştirirlerdi ve bu, bütün

uygulamanın amacını yerle bir ederdi. Dahası, burada durum bu olmasa da değerlendirmelerin tanımında beklenmeyen belirleyici düşüncelerin ortaya çıkması olasılığına karşı temkini elden bırakmayacaktı (Bölüm 10'da bahsi geçen kırık bacak faktörleri). Eğer bu tür umulmadık anlaşma bozucular (ya da tam tersine, anlaşmayı güçlendirici faktörler) ortaya çıkarsa, değerlendirmelerin ortalamasına dayanan tamamen mekanik bir süreç ciddi bir hataya yol açabilir.

Joan ayrıca bu aşamada yönetim kurulu üyelerinin sezgilerini kullanmalarına izin vermenin, sürecin daha önceki kısımlarında izin vermekten çok farklı olduğunu biliyordu. Artık değerlendirmeler herkes tarafından ulaşılabilir ve bilinir olduğu için, nihai karar güvenli bir şekilde bu olguya dayalı, detaylı olarak tartışılmış puanlamalara dayanacaktı. Herhangi bir yönetim kurulu üyesinin, karşısında çoğu destekleyici maddelerden oluşan bir aracı değerlendirme listesi dururken, anlaşmaya karşı çıkabilmesi için çok güçlü nedenler sunması gerekirdi. Bu mantığa uyarak, kurul anlaşmayı, neredeyse bütün kurulların yaptığı gibi, tartıştı ve oyladı.

Tekrarlı Kararlarda Aracı Değerlendirmeler Protokolü

Aracı değerlendirmeler protokolünü tek seferlik, tekil kararlar bağlamında tanımladık. Ancak prosedür tekrarlı kararlara da uygulanabilir. Mapco'nun tek seferlik bir şirket alımı yapmadığını ama yeni kurulan şirketlere tekrarlı yatırımlar yapan bir risk sermayesi şirketi olduğunu hayal edin. Protokol, durumu daha basitleştirecek iki nokta dışında yine aynı şekilde uygulanabilir, hikâye neredeyse aynı olurdu.

İlk olarak, birinci adımın –aracı değerlendirmeler listesini tanımlama işinin– sadece bir kez yapılması gerekir. Şirketin bütün potansiyel yatırımlarına uygulanan yatırım kriterleri zaten

var: Değerlendirmeler bu kriterler olurdu. Onları her defasında yeniden keşfetmek gerekmezdi.

İkinci olarak, eğer şirket aynı türden çok fazla karar alıyorsa, tecrübesini yargılarını kalibre etmek için kullanabilir. Örneğin, her şirketin yapmak isteyeceği bir değerlendirme düşünün: yönetim ekibinin kalitesinin değerlendirilmesi. Bu tür değerlendirmelerin bir referans sınıfına göre yapılması gerektiğini söylemiştik. Belki Mapco'nun analizcilerinin hâlinden anlamışsınızdır: Belirli bir hedefi değerlendirmeye ek olarak karşılaştırılabilir şirketler hakkında bilgi toplamak zorlu bir iştir.

Karşılaştırmalı yargılar tekrarlı bir karar bağlamında kolaylaşırlar. Düzinelerce hatta yüzlerce şirketin yönetim ekibini değerlendirdiyseniz, bu ortak tecrübeyi bir referans sınıfı olarak kullanabilirsiniz. Bunu yapmanın pratik bir yolu, çapa vakalarla tanımlanmış bir vaka ölçeği yaratmaktır. Örneğin, hedef yönetim ekibinin "ABC şirketinin bizim aldığımız zamanki yönetim ekibi kadar iyi olduğunu" ama "DEF şirketinin yönetim ekibi kadar iyi olmadığını" söyleyebilirsiniz. Çapa vakalar, elbette bütün katılımcılar tarafından biliniyor (ve periyodik olarak güncelleniyor) olmalıdır. Onları tanımlamak önceden zaman yatırımı yapmayı gerektirir. Ama bu yaklaşımın değeri, göreceli yargıların (bu ekibi ABC ve DEF'tekilerle karşılaştırmak) sayılar ya da sıfatlarla tanımlanmış bir ölçekteki mutlak puanlamalardan daha güvenilirdir.

Protokolün Değiştirdikleri

Referans kolaylığı sağlamak için, aracı değerlendirmeler protokolünün gerektirdiği ana değişiklikleri Tablo 4'te özetledik.

Tablo 4: *Aracı değerlendirmeler protokolünün ana adımları*

1. Sürecin başlangıcında, kararı aracı değerlendirmeler şeklinde yapılandırın. (*Tekrarlı yargılar için bu sadece bir kez yapılır.*)
2. Mümkün olan her durumda, aracı değerlendirmelerin bir dış bakış kullandığından emin olun. (*Tekrarlı yargılar için: mümkünse bir vaka ölçeğiyle, göreceli yargılar kullanın.*)
3. Analitik aşamada, değerlendirmeleri birbirlerinden olabildiğince bağımsız tutun.
4. Karar toplantısında, her değerlendirmeyi ayrı ayrı gözden geçirin.
5. Her değerlendirmede, katılımcıların yargılarına ayrı ayrı vardıklarından emin oldun, sonra tahmin-konuşma-tahmin yöntemini kullanın.
6. Nihai kararı vermek için sezgiyi geciktirin ama yasaklamayın.

Burada, bir önceki bölümde sunduğumuz karar hijyeni tekniklerinin birçoğunun uygulandığını fark etmiş olabilirsiniz: bilgiyi sıralama, kararı bağımsız yargılara yapılandırma, dış bakışa dayanan ortak bir referans çerçevesi kullanma, birden fazla bireyin bağımsız yargılarını bir araya toplama. Aracı değerlendirmeler protokolü, bu teknikleri uygulayarak, karar *sürecini* olabildiğince çok karar hijyeni içerecek şekilde değiştirmeyi hedefler.

Kararların içeriğine karşı sürecin vurgulanması, şüpheyle karşılanabilir. Araştırma ekibi üyelerinin ve yönetim kurulu üyelerinin tepkileri, anlattığımız hâlleriyle, sıra dışı değil. İçerik duruma özel, süreç ise genele uygundur. Sezgi ve yargı kullanmak eğlencelidir ama bir sürece uymak, eğlenceli değildir. Geleneksel bilgelik iyi kararların –özellikle en iyilerinin– büyük liderlerin içgörüsü ve yaratıcılığından doğduğunu savunur. (Söz konusu li-

der biz olduğumuzda buna inanmak özellikle hoşumuza gider.) Ve birçok insan için *süreç* kelimesi bürokrasiyi, kırtasiye işlerini ve gecikmeleri çağrıştırır.

Protokolün ögelerinin tamamını ya da bir kısmını uygulamaya koyan şirketler ve hükûmet bürolarıyla tecrübelerimiz bu endişelerin yanlış yönlendirilmiş olabileceğini düşündürmektedir. Hiç şüphesiz, zaten bürokratik olan bir örgütün karar alma süreçlerine karmaşıklık eklemek durumu iyileştirmez. Ama karar hijyeninin yavaş ya da bürokratik olması şart değildir. Aksine, karar hijyeni, bürokrasilerin en önemli özelliği olan boğucu konsensüsü değil, meydan okumayı ve tartışmayı destekler.

Karar hijyeninin artıları çok nettir. İş dünyasındaki ve kamu sektöründeki liderler genellikle en büyük ve en önemli kararlarındaki gürültüden tamamen bihaberdirler. Sonuç olarak gürültüyü azaltmak için özel bir önlem almazlar. Bu açıdan, tek personel seçme aracı olarak yapılandırılmamış mülakatları kullanmaya devam eden işe alım uzmanlarından bir farkları yoktur. Kendi yargılarının gürültüsünden bihaber, yargılarının geçerliliğinden olması gerekenden çok daha emin ve yargılarını iyileştirebilecek prosedürlerden habersizdirler.

El yıkamak bütün hastalıkları önlemez. Aynı şekilde karar hijyeni de bütün hataları önlemeyecektir. Bütün kararları muhteşem de kılmayacaktır. Ama el yıkamak gibi, gözle görülmez ama yaygın ve hasar verici bir sorunu ele alır. Yargının olduğu yerde gürültü vardır ve karar hijyenini gürültüyü azaltmak için bir araç olarak sunuyoruz.

Aracı Değerlendirmeler Protokolünden Bahsederken

"İşe alım kararları için yapılandırılmış bir sürecimiz var. Stratejik kararlar için neden olmasın? Sonuçta seçenekler de adaylara benzer."

"Bu zor bir karar. Üstüne kurulması gereken aracı değerlendirmeler neler?"

"Bu plan hakkında sezgisel, bütüncül yargımız çok önemli ama onu henüz tartışmayalım. Sezgimiz, istediğimiz ayrı değerlendirmelerden destek aldığı zaman çok daha fazla işe yarayacaktır."

KISIM 6

Optimal Gürültü

1973'te Yargıç Marvin Frankel ceza hükümlerinde gürültüyü azaltma çabası için sürekli bir çaba çağrısı yaparken çok haklıydı. Frankel'in gayriresmî ve sezgisel gürültü denetimi ve onu takip eden daha resmî ve sistemli çabalar, benzer insanların gördüğü muamelelerde gerekçesiz uyuşmazlıkları ortaya çıkardı. Bu uyuşmazlıklar hem korkunç hem şaşırtıcıydılar.

Bu kitabın büyük kısmı Frankel'in argümanlarını genelleştirme ve psikolojik temellerine yönelik bir anlayış sunma çabası olarak anlaşılabilir. Bazı insanlar için, ceza yargılama sistemindeki gürültü benzersiz bir şekilde katlanılmaz bir şey, hatta kepazelik olabilir. Ancak sayısız diğer bağlamda da özel ve kamu sektörlerinde birbirinin yerine geçebilen farklı insanların iş hakkında farklı kararlar alması nedeniyle, tam olarak katlanılır olduğu söylenemez. Sigortacılık, çalışanların işe alımı ve değerlendirilmesi, tıp, adli tıp, eğitim, iş dünyası ve hükûmette, kişiler arası gürültü başlıca hata kaynaklarından biridir. Ayrıca, her birimizin, sözde alakasız faktörlerin bizi sabah ve öğleden sonra ya da pazartesi ve perşembe günleri farklı yargılara yönlendirebilmesi anlamında, durum gürültüsüne açık olduğumuzu da gördük.

Ancak cezai hüküm kılavuzlarına gösterilen yoğun yargısal tepkinin düşündürdüğü gibi, gürültüyü azaltma çabaları sık sık ciddi, hatta tutkulu itirazlarla karşı karşıya kalmaktadır. Pek çok insan kılavuzların çok katı, insanlık dışı ve kendilerince adaletsiz olduğunu savundu. Neredeyse hepimizin bir şirketten, bir çalışandan ya da bir hükûmetten makul bir talepte bulunup karşılığında "Size yardımcı olmayı çok isterdik. Ama elimiz kolumuz bağlı. Burada çok açık kurallarımız var." tepkisini aldığımız en az bir tecrübemiz olmuştur. Söz konusu kurallar aptalca ve hatta acımasızca görünebilir ama benimsenmelerinin iyi bir nedeni vardır: gürültüyü (ve belki de ön yargıyı) azaltmak.

Buna rağmen, bazı gürültüyü azaltma çabaları, özellikle insanların adil bir şekilde dinlenmesini zorlaştırmaları ya da imkânsız kılmaları durumunda, ciddi endişelere yol açar. Algoritmaların ve yapay öğrenmenin kullanılması bu amaca bambaşka bir ışık tutmuştur. Hiç kimse, "Algoritmaların zamanı geldi!" diyen bir pankartın altında yürümez.

Etkili bir eleştiri, Yale Hukuk Fakültesi'nden Kate Stith ve federal yargıç José Cabranes'ten geldi. Cezai hüküm kılavuzlarına ve bir anlamda, buradaki merkez argümanlarımızdan birine hararetli bir saldırıları oldu. Argümanları cezai hüküm alanıyla sınırlıydı ama aynı zamanda eğitim, iş dünyası, spor ve diğer her yerdeki pek çok gürültüyü azaltma stratejisine bir itiraz olarak kabul edilebilirdi. Stith ve Cabranes cezai hüküm kılavuzlarının "takdir yetkisinin kullanılması korkusu –yargılama korkusu– uzmanlara ve merkezî planlamaya teknokratik bağlılık" ile canlandığını öne sürüyorlar. "Yargılama korkusunun" "her vakanın kendine has özelliklerinin" değerlendirilmesini menettiğini savunuyorlar. Onların bakış açısıyla, "hiçbir mekanik çözüm adaletin taleplerini karşılayamaz."

Bu itirazlar incelenmeye değer. Her tür yargıyı barındıran ortamlarda, insanlar "adaletin taleplerini" her tür mekanik çözümü yasaklayan ve dolayısıyla gürültüyü garantileyen süreç ve

yaklaşımlara izin veren ve hatta onları zorunlu kılan bir şey olarak görüyorlar. Birçok insan, "her vakanın kendine has özelliklerine" dikkat çekiyor. Hastanelerde, okullarda ve büyük ya da küçük şirketlerde, bu çağrının derin sezgisel bir cazibesi var. Karar hijyeninin gürültüyü azaltmak için çeşitli stratejiler içerdiğini ve bu stratejilerin çoğunun mekanik çözümler barındırmadığını gördük: İnsanlar bir sorunu ögelerine parçaladıkları zaman, yargılarının mekanik olması gerekmez. Buna rağmen, pek çok insan karar hijyeni stratejilerinin kullanılmasını hoş karşılamıyor.

Gürültüyü istenmeyen bir değişkenlik olarak tanımladık ve eğer bir şey istenmiyorsa, muhtemelen bertaraf edilmelidir. Ama bu analiz bundan çok daha karmaşık ve daha ilginç. Gürültü, diğer şeyler eşit olduğunda, istenmeyebilir. Ama diğer şeyler eşit olmayabilir ve gürültüyü ortadan kaldırmanın maliyeti getirilerini aşabilir. Ve bir maliyet ve getiri analizi gürültünün maliyetli olduğunu düşündürdüğü zaman bile, onu ortadan kaldırmak hem kamusal hem özel kurumlar için bir dizi korkunç ve hatta kabul edilemez sonuç doğurabilir.

Gürültüyü azaltma ya da ortadan kaldırma çabalarına başlıca yedi itiraz vardır.

Öncelikle, gürültüyü azaltmak pahalı olabilir; zahmete değmeyebilir. Gürültüyü azaltmak için gerekli adımlar çok zahmetli olabilir. Bazı durumlarda uygulanabilir bile olmayabilirler.

İkinci olarak gürültüyü azaltmak için devreye sokulan bazı stratejiler kendi hatalarını da yanlarında getirebilirler. Zaman zaman, sistematik ön yargı üretebilirler. Bir hükûmet bürosundaki bütün tahminciler aynı gerçek dışı iyimser varsayımları benimserse, tahminleri gürültülü olmaz ama yanlış olur. Bir hastanedeki bütün doktorlar her hastalık için aspirin reçete etse, gürültülü olmazlar ama birçok hata yapmış olurlar.

Bu itirazları Bölüm 26'da keşfedeceğiz. Bölüm 27'de, yine yaygın olan ve önümüzdeki yıllarda, özellikle kurallara, algoritmalara

ve yapay öğrenmeye gittikçe daha fazla bel bağlanması sonucu, pek çok yerde duyulması muhtemel diğer beş itiraza değineceğiz.

Üçüncü olarak, insanların saygılı muamele gördüklerini hissetmelerini istiyorsak bir miktar gürültüyü hoş görmek zorunda kalabiliriz. Gürültü, nihayetinde benimsenmek zorunda kalınan kusurlu süreçlerin bir yan ürünü olabilir çünkü söz konusu süreç herkese (çalışanlara, müşterilere, başvuru sahiplerine, öğrencilere ve bir suçla suçlanan kişilere) bireyselleşmiş bir kendini ifade fırsatı, takdir yetkisinin kullanımını etkileme imkânı ve görülmek ve duyulmak için bir şanslarının olduğu hissini verir.

Dördüncü olarak, gürültü yeni değerleri barındırmak ve dolayısıyla ahlaki ve politik gelişime olanak sağlamak için gerekli olabilir. Gürültüyü ortadan kaldırırsak, ahlaki ve politik taahhütler yeni ve beklenmedik yönlere kaydığı zaman tepki verme becerimizi de azaltmış olabiliriz. Gürültüden tamamen arınmış bir sistem, mevcut değerleri dondurabilir.

Beşinci olarak, gürültüyü azaltmak için tasarlanmış bazı stratejiler fırsatçı davranışı teşvik ederek, insanların sistemi manipüle etmesine ya da yasaklamalardan kaçınabilmesine olanak sağlayabilir. Hata yapılmasını önlemek için biraz gürültü ya da belki de çokça gürültü gerekli olabilir.

Altıncı olarak, gürültülü bir süreç iyi bir caydırıcı olabilir. İnsanlar küçük ya da büyük bir cezaya maruz kalabileceklerini bilirlerse, bir kabahat işlemekten uzak durabilirler; en azından riskten hoşlanmıyorlarsa. Sistemler ekstra caydırıcılık üretme yolu olarak gürültüyü hoş görebilir.

Son olarak, insanlar onlara sadece nesne ya da bir makinenin dişlileri gibi davranılmasını istemez. Kimi gürültüyü azaltma stratejileri, insanların yaratıcılığını ezdiği için moral bozucu olabilir.

Bu itirazları elimizden geldiğince sempatik bir şekilde ele alsak da hiçbir şekilde desteklemediğimizi belirtmek isteriz. En azından genel gürültüyü azaltma amacını reddetme nedenleri ola-

rak kullanılıyorlarsa. Kendini tekrarlayacak bir hususu önceden belirtmek için, bir itirazın ikna edici olup olmaması hedef aldığı gürültüyü azaltma stratejisine bağlıdır. Örneğin, katı kılavuzlara itiraz ederken, bağımsız yargıların bir araya toplanmasının iyi bir fikir olduğu konusunda hemfikir olabilirsiniz. Aracı değerlendirmeler protokolünü kullanmaya itiraz ederken, dış bakış üstüne kurulu ortak bir ölçeğin kullanımını şiddetle destekleyebilirsiniz. Bu noktaları aklımızda tutarak, genel çıkarımımız itirazlara hak verildiğinde bile, gürültünün azaltılmasının uğraşılmaya değer ve hatta acil bir hedef olmaya devam ettiğidir. Bölüm 28'de bu çıkarımı insanların her zaman farkında olmasalar bile, her gün karşılaştığı bir ikilemi inceleyerek savunacağız.

BÖLÜM 26

Gürültüyü Azaltmanın Maliyetleri

İnsanlardan gürültüyü ortadan kaldırmaları istendiğinde, gerekli adımların çok pahalı olduğunu söyleyerek itiraz edebilirler. Abartılı koşullarda, gürültüyü azaltmak açıkça mümkün değildir. Bu itirazı iş dünyasında, eğitimde, hükûmette ve başka yerlerde duyduk. Burada meşru bir endişe var ama bu endişe kolayca abartılabilir ve genellikle sadece bir bahanedir.

İtirazı en cazip şekle sokmak için, okul yılı boyunca her hafta, onuncu sınıf öğrencilerinin yazdığı yirmi beş kompozisyona not veren bir lise öğretmenini düşünün. Öğretmen her kompozisyona on beş dakikadan daha kısa zaman ayırırsa, not verme süreci gürültülü ve dolayısıyla hatalı ve haksız olabilir. Öğretmen küçük bir karar hijyenini uygulamayı düşünerek, gürültüyü azaltmak için bir meslektaşından da not vermesini rica edebilir. Böylece her kâğıdı iki kişi okumuş olur. Belki de öğretmen aynı amaca, her kompozisyona daha fazla zaman ayırarak, göreceli olarak karmaşık değerlendirme sürecini yapılandırarak ya da kompozisyonları birden fazla kez ve farklı sırayla okuyarak ulaşabilir. Kontrol lis-

tesi olarak kullanılan detaylı bir not verme kılavuzu da yardımcı olabilir. Ya da eğitimci durum gürültüsünü azaltmak için bütün kompozisyonları günün aynı saatinde okumaya gayret edebilir.

Ancak öğretmenin kendi yargıları hayli isabetli ise ve korkunç derecede gürültülü değilse, bunların hiçbirini yapmamak mantıklı olabilir. O zahmete girmeye değmeyebilir. Öğretmen bir kontrol listesi kullanmanın ya da bir meslektaşından aynı kâğıtları okumasını rica etmenin bir tür aşırılık olacağını düşünebilir. Öyle olup olmadığını anlamak için, disiplinli bir analiz gerekebilir: Öğretmen daha ne kadar isabetlilik kazanır, daha fazla isabetlilik ne kadar önemlidir, gürültüyü azaltma çabası için ne kadar zaman ve para gerekir? Gürültüyü azaltmaya ne kadar yatırım yapılacağı konusunda kolayca bir sınır belirleyebiliriz. Kompozisyonların dokuzuncu sınıf öğrencileri tarafından ya da üniversiteye kabulün kapıda ve riskin daha büyük olduğu son sınıf tezi olarak yazıldığı durumlarda sınırın farklı olması gerektiğini de yine aynı kolaylıkla görebiliriz.

Temel analiz, her tür özel ve kamusal örgütün karşılaştığı ve onları bazı gürültüyü azaltma stratejilerini reddetmeye yönlendiren daha karmaşık durumlara genişletilebilir. Bazı hastalıklarda, hastaneler ve doktorlar çeşitliliği ortadan kaldırmak için basit kılavuzlar belirlemek için mücadele verebilirler. Uyumsuz tıbbi teşhisler konusunda, gürültüyü azaltma çabalarının özel bir cazibesi vardır: Hayat kurtarabilirler. Ancak bu çabaların uygulanabilirliği ve maliyeti hesaba katılmalıdır. Bir test teşhislerde gürültüyü azaltabilir ama eğer söz konusu test kullanışsız, tehlikeli ve maliyetliyse ve teşhislerdeki değişkenlik çok azsa ve çok hafif sonuçlara neden oluyorsa, bütün doktorların bütün hastalardan bu teste girmelerini istemelerine değmeyebilir.

Çalışanların değerlendirilmesinin ölüm kalım meselesi olması çok nadirdir. Ama gürültü çalışanlar için haksızlığa ve şirket için yüksek maliyetlere neden olabilir. Gürültüyü azaltma çabalarının

uygulanabilir olması gerektiğini gördük. Buna değiyorlar mı? Açıkça hatalı değerlendirmeleri içeren vakalar fark edilebilir ve utanç verici ya da daha beter olarak düşünülebilir. Bununla birlikte, kurumlar detaylı düzeltici adımların harcanan çabaya değmeyeceğini de düşünebilirler. Bazen bu çıkarım, uzağı göremeyen, kişinin ya da kurumun kendisine hizmet eden ve hatalı –hatta feci derecede hatalı– bir çıkarım olabilir. Bazı karar hijyeni türleri çabaya değebilir. Ancak gürültünün azaltılmasının çok pahalı olduğu inancı her zaman yanlış değildir.

Kısaca, gürültüyü azaltmanın getirilerini maliyetle karşılaştırmalıyız. Bu yerinde bir tavır olur ve gürültü denetimlerinin çok önemli olmasının nedenlerinden biridir. Pek çok durumda, denetimler gürültünün korkunç haksızlık düzeyleri, çok yüksek maliyetler ya da her ikisini birden ürettiğini ortaya çıkarabilir. Bu durumda, gürültüyü azaltmanın maliyeti çabayı harcamamak için iyi bir neden olmaktan çıkacaktır.

Daha Az Gürültü, Daha Fazla Hata?

Farklı bir itiraz da gürültüyü azaltma çabalarının kendilerinin kabul edilemeyecek düzeyde hatalar üretebilmesidir. İtiraz, gürültüyü azaltmak için kullanılan aletler fazla kör olduğunda, ikna edici olabilir. Aslında, bazı gürültüyü azaltma çabaları ön yargıyı artırabilir. Facebook ya da Twitter gibi bir sosyal medya platformu belli kaba sözcükleri içeren bütün paylaşımları kaldırmayı gerektiren sıkı kurallar uygulamaya başlasa, kararları daha az gürültülü olur ama bu kalmasına izin verilmesi gereken çok fazla paylaşımın da kaldırılması anlamına gelir. Bu sahte olumlular bir yönsel hata, bir ön yargıdır.

Hayat insanların takdir yetkisini kullanmasını ve gürültü yaratan uygulamaları azaltmak için tasarlanmış kurumsal reformlarla doludur. Bu tür reformların çoğu iyi motive edilmiştir

ancak bazı tedaviler hastalığın kendisinden beterdir. Ekonomist Albert Hirschman, *The Rhetoric of Reaction* adlı kitabında, reform çabalarına üç yaygın itiraza dikkat çeker. Öncelikle, bu tür çabalar çözmeyi amaçladıkları sorunu ağırlaştırmaları açısından ters etki yaratabilirler. İkinci olarak, işe yaramaz olabilirler; hiçbir şeyi değiştirmeyebilirler. Üçüncü olarak diğer önemli değerleri tehlikeye atabilirler (Örneğin iş sendikalarını ve sendikalaşma hakkını koruma çabasının ekonomik büyümeye zarar verdiğinin söylenmesi gibi). Ters etki, işe yaramama ve tehlikeye atma gürültüyü azaltmaya itiraz olarak öne sürülebilir ve üçü arasında ters etki ve tehlikeye atma en güçlüler olarak görünürler. Bazen bu itirazlar sadece retoriktir; aslında çok iyilik üretecek bir reformu yolundan çıkarma çabasıdır. Ancak, bazı gürültüyü azaltma stratejileri önemli değerleri tehlikeye atabilir ve diğerlerinde ters etki riski kolaylıkla geçiştirilemeyebilir.

Cezai hüküm kılavuzlarına itiraz eden yargıçlar da bu riske dikkat çekiyordu. Yargıç Frankel'in çalışmasından haberdardılar ve takdir yetkisinin gürültüye neden olduğunu inkâr etmiyorlardı. Ancak takdir yetkisini azaltmanın daha az değil, daha çok soruna neden olacağını düşünüyorlardı. Václav Havel'den bir alıntıyla, "Dünyanın sadece çözülmesi gereken bir yapboz, kullanım talimatları olan, keşfedilmeyi bekleyen bir makine, önünde sonunda evrensel bir çözüm üreteceği umuduyla bir bilgisayara yüklenen bir bilgi kütlesi olduğu inancını terk etmeliyiz." Evrensel çözümler fikrini reddetmenin bir nedeni insana özgü durumların hayli çeşitli olduğuna ve iyi yargıçların bu değişkenlikleri ele aldığına duyulan ısrarcı inançtır ki bu gürültüyü hoş görmek ya da en azından birtakım gürültüyü azaltıcı stratejileri reddetmek anlamına gelebilir.

Bilgisayar satrancının ilk günlerinde, büyük bir hava yolu şirketi uluslararası yolcularına bilgisayara karşı oynamaya davet edildikleri bir satranç programı sunuyordu. Programın birkaç seviyesi vardı. En düşük seviyede, program basit bir kural kullanıyordu: Her

fırsatta rakibinizin şahını sıkıştırın. Program gürültülü değildi. Her defasında aynı şekilde oynuyordu, her zaman basit kuralına uyuyordu. Ama kural çokça hatayı garantiliyordu. Program satrançta çok kötüydü. Tecrübesiz bir satranç oyuncusu bile onu yenebilirdi (hiç şüphesiz amaç buydu; kazanan uçak yolcuları, mutlu uçak yolcuları olur).

Ya da Amerika'nın bazı eyaletlerinde benimsenen ve "üç vuruş ve oyun dışısınız" olarak adlandırılan cezai hüküm politikasını düşünün. Fikir şu: Üç suç işlediğiniz zaman, cezanız müebbet hapis. Nokta. Bu politika hükmü verecek yargıcın rastgele görevlendirilmesinden kaynaklanan değişkenliği azaltır. Savunucularından bazıları, yatay gürültüden ve bazı yargıçların gedikli suçlulara karşı fazla merhametli olmasından özellikle endişeliydiler. Gürültüyü ortadan kaldırmak, üç vuruş mevzuatının merkez noktasıdır.

Ancak üç vuruş politikası gürültüyü azaltma amacında başarılı olsa bile, bu başarının bedelinin çok yüksek olduğunu söyleyerek makul bir itirazda bulunabiliriz. Üç suç işlemiş bazı insanlar ömür boyu hapse kapatılmamalıdır. Belki suçları şiddet içermiyordur. Ya da korkunç hayat şartları onları suça yönlendirmiş olabilir. Belki de rehabilitasyona yatkındırlar. Pek çok insan müebbet hapis cezasının, belli şartlara dikkat etmeden, sadece çok sert değil, aynı zamanda dayanılmayacak kadar katı olduğunu düşünür. Bu nedenle, bu gürültüyü azaltma stratejisinin bedeli çok yüksektir.

ABD Yüksek Mahkemesi'nin zorunlu idam cezasının çok gaddarca olduğu için değil, *kural olduğu için* anayasaya aykırı olduğuna hükmettiği *Woodson-Kuzey Carolina Eyaleti* davasını düşünün. Zorunlu idam cezasının bütün amacı gürültüyü önlemek, belirtilen şartlar altında, katillerin öldürülmesi gerektiğini söylemekti. Bireyselleştirilmiş muamele ihtiyacını dile getiren mahkeme, "benzer yasal kategorideki her suçun belirli bir suçlunun geçmiş hayatına ya da alışkanlıklarına bakılmaksızın aynı cezayı gerektirdiği inancı artık geçerli değildir." demişti. Yüksek

Mahkeme'ye göre, zorunlu ölüm cezasının ciddi anayasal eksiği, "belirli bir suçtan hüküm giyen herkese benzersiz bireyler olarak değil, körü körüne ölüm cezasına çarptırılacak, sureti olmayan, birbirinden ayırt edilemez bir kütle olarak muamele etmesidir."

Elbette, ölüm cezasında ortada özellikle yüksek bir risk ve tehlike vardır ancak mahkemenin analizi çoğu hukukla ilgili olmayan daha pek çok duruma uyarlanabilir. Öğrencileri değerlendiren öğretmenler, hastaları değerlendiren doktorlar, çalışanları değerlendiren işverenler, sigorta primi belirleyen risk değerlendirme uzmanları, sporcuları değerlendiren koçlar, bu insanların hepsi aşırı katı, gürültüyü azaltıcı kurallar uygularlarsa hata yapabilirler. İşverenler çalışanları değerlendirme, terfi ya da askıya alma konusunda basit kurallar kullanırlarsa, bu kurallar gürültüyü ortadan kaldırırken, çalışanların performansının önemli yönlerini de yok sayabilirler. Önemli değişkenleri hesaba katmayan gürültüden arınmış bir puanlama sistemi (gürültülü) bireysel yargılara bel bağlamaktan daha kötü olabilir.

Bölüm 27'de insanlara "sureti olmayan, birbirinden ayırt edilemez bir kütle" yerine, "benzersiz bireyler" olarak muamele etme fikrini ele alacağız. Şimdilik, daha sıkıcı bir noktaya odaklanalım. Bazı gürültüyü azaltma stratejileri çok fazla hataya neden olur. Şu aptal satranç programına çok benzeyebilirler.

Yine de, itiraz aslında olduğundan çok daha ikna edici görünmektedir. Bir gürültüyü azaltma stratejisi hataya yatkınsa, yüksek gürültü seviyelerine razı olmamalıyız. Bunun yerine, daha iyi bir gürültüyü azaltma stratejisi geliştirmeye çalışmalıyız. Örneğin, saçma kurallar benimsemek yerine yargıları bir araya toplayarak ya da aptalca kılavuzlar ve kurallar yerine, akıllıca olanları geliştirerek. Gürültüyü azaltma yararına, bir üniversite, en yüksek test sonuçlarının kabul edileceğini söyleyebilir, bu kadar. Bu kural fazla ham gelirse, okul test sonuçlarını, notları, yaşı, atletik başarıları, aile geçmişini ve daha fazlasını hesaba katabilecek bir

formül yaratabilir. Karmaşık kurallar daha isabetli, ilgili faktörlerin tam aralığına daha uyumlu olabilir. Benzer bir şekilde, doktorların bazı hastalıkları teşhis etmek için karmaşık kuralları vardır. Profesyoneller tarafından kullanılan kılavuzlar ve kurallar her zaman basit ya da ham değildir ve çoğu gürültüyü dayanılmayacak kadar yüksek maliyetler (ya da ön yargı) yaratmadan azaltmaya yardımcı olabilirler. Ve kılavuzlar ya da kurallar işe yaramazsa, belki özellikle o duruma uygun düşen, yargıları bir araya toplamayı ya da aracı değerlendirmeler protokolü gibi yapılandırılmış bir süreci kullanmayı geri çağıran diğer karar hijyeni türlerini devreye sokabiliriz.

Gürültüsüz, Ön Yargılı Algoritmalar

Gürültüyü azaltmanın potansiyel olarak yüksek maliyetleri genellikle "algoritma ön yargısına" itirazların gittikçe arttığı algoritmalar bağlamında karşımıza çıkar. Daha önce gördüğümüz gibi, algoritmalar gürültüyü ortadan kaldırır ve genellikle bu nedenden cazip görünür. Hatta bu kitabın büyük kısmı sadece daha az gürültülü oldukları için, algoritmalara daha fazla yer ayrılmasının savunması olarak görülebilir. Ancak gördüğümüz gibi, algoritmalara daha fazla bel bağlanması ırk ve cinsiyet bazında ya da dezavantajlı grupların üyelerinin aleyhine ayrımcılığı artırıyorsa, gürültüyü azaltmanın maliyeti katlanılmayacak kadar büyük olabilir.

Algoritmaların bu ayrımcı sonucu yaratacağı korkusu çok yaygındır. Hiç şüphesiz bu çok ciddi bir risk olur. Matematikçi Cathy O'Neil, *Weapons of Math Destruction*'da büyük verilere ve algoritma kararlarına bel bağlamanın, peşin hüküm içerebileceğini, eşitsizliği artırabileceğini ve demokrasiyi tehdit edebileceğini öne sürer. Bir diğer şüpheci yaklaşıma göre, "Potansiyel olarak ön yargılı matematik modelleri hayatımızı yeniden düzenliyor ve ne geliştirilmelerinden sorumlu şirketler ne de hükûmet bu sorunu

gidermekle ilgileniyor." Bağımsız bir sorgulamacı gazetecilik örgütü olan ProPublica'ya göre, suçun tekrarlanması risk değerlendirmesinde yaygın olarak kullanılan COMPAS adlı algoritma ırksal azınlıkların aleyhine güçlü ön yargılar sergiliyor.

Gürültüden arınmış ama aynı zamanda ırkçı, cinsiyetçi ya da farklı şekilde ön yargılı bir algoritma yaratmanın mümkün – hatta kolay– olduğundan kimse şüphe duymamalı. O kişiye kefalet hakkının tanınıp tanınmayacağını belirlemek için, aleni olarak sanığın ten rengini kullanan bir algoritma ayrımcıdır (ve kullanımı birçok ülkede kanuna aykırı olacaktır). İş başvurusu yapan adayların hamile kalma olasılığını hesaba katan bir algoritma kadınlara karşı ayrımcıdır. Bu ve benzer durumlarda, algoritmalar yargıda istenmeyen değişkenliği bertaraf edebilir ama kabul edilemez ön yargılar barındırır.

Prensipte, ırkı ya da cinsiyeti hesaba katmayan bir algoritma tasarlayabilmeliyiz. Hatta ırkı ve cinsiyeti tamamen görmezden gelen bir algoritma da tasarlanabilir. Daha zorlayıcı olan ve artık çok dikkat çeken sorun, bir algoritmanın ırkı ve cinsiyeti açıkça kestirici olarak kullanmadığında bile, ayrımcılık yapabileceği ve bu anlamda, ön yargılı olabileceğidir.

Daha önce belirttiğimiz gibi, bir algoritma iki ana nedenden ön yargılı olabilir. Öncellikle, tasarım gereği ya da tasarım gereği olmadan, ırk ve cinsiyetle yüksek korelasyonu olan kestiriciler kullanabilir. Örneğin kilo ve boy cinsiyetle bağıntılıdır ve insanların büyüdüğü ya da yaşadığı yer, pekâlâ ırkla bağıntılı olabilir. İkinci olarak ayrımcılık kaynak veriden de gelebilir. Eğer bir algoritma ön yargılı bir veri setiyle eğitilmişse, kendisi de ön yargılı olacaktır. Genellikle, polis kaynaklarının tahsisini iyileştirmek için, suçu öngörmeye çalışan "tahmine dayalı polislik" algoritmalarını düşünün. Suçla ilgili mevcut veri belli mahallelerde normalin üstünde polis faaliyeti olduğunu ya da belirli suç çeşitlerinin göreceli olarak daha fazla raporlandığını yansıtırsa, o zaman sonucu

olan algoritmalar ayrımcılığı sürdürecek ya da alevlendirecektir. Beslenen veride ön yargı olduğunda, isteyerek ya da istemeyerek, ayrımcılığı kodlayan bir algoritma tasarlamak hayli mümkündür. Bir algoritma ırkı ve cinsiyeti kasten hesaba katmasa bile, tıpkı insanlar gibi ön yargılı olabileceği görülmüştür. Hatta bu açıdan, algoritmalar daha kötü olabilir; gürültüyü ortadan kaldırdıkları için insan yargıçlardan daha *güvenilir şekilde* ön yargılı olabilirler.

Pek çok insan için kilit pratik bir değerlendirme, bir algoritmanın ayırt edilebilir gruplar üstünde farklı etkisi olup olmadığıdır. Farklı etkinin tam olarak nasıl test edileceği ve bir algoritma için ayrımcılığı, ön yargıyı ya da hakkaniyeti nelerin oluşturduğuna karar verilmesi, bu kitabın kapsamını aşan şaşırtıcı derecede karmaşık konulardır.

Bu sorunun gündeme gelebilmesi ise algoritmaların insan yargıları karşısındaki belirgin bir avantajıdır. Öncelikle, kabul edilemez girdileri değerlendirmeye almadıklarından emin olmak ve itiraz edilebilir bir şekilde ayrımcılık yapıp yapmadıklarını test etmek için, algoritmaların dikkatle değerlendirilmesini öneriyoruz. Yargıları genellikle opak olan insanları aynı türde bir incelemeye tabi tutmak çok daha zordur; insanlar bazen bilinçsiz olarak ve yasal sistem dâhil dış gözlemcilerin kolayca göremeyeceği şekilde ayrımcılık yaparlar. Bu nedenle, bazı açılardan algoritmalar insanlardan daha şeffaf olabilir.

Hiç şüphesiz, gürültüsüz ama ön yargılı algoritmalara dikkat çekmeliyiz; tıpkı gürültüsüz ama ön yargıları olan kuralların maliyetlerini hesaba katmamız gerektiği gibi. Kilit soru önemli olan bir kriterler kombinasyonu üstünde gerçek dünyadaki insan karar alıcılardan daha iyi iş çıkaran algoritmalar tasarlayıp tasarlayamayacağımızdır. Algoritmaların hangi kriter kombinasyonunu seçersek seçelim, insanlara üstünlük sağlayabileceğini gösteren kanıtların miktarı hiç az değildir. (Sağlayacağını değil, sağlayabileceğini dediğimize dikkatinizi çekmek isteriz.) Örneğin Bölüm

10'da bahsedildiği gibi, bir algoritma kefalet kararları konusunda insan yargıçlardan daha isabetli olurken, insanlardan daha az ırksal ayrımcılık üretebilir. Benzer bir şekilde, bir öz geçmiş seçme algoritması, insan öz geçmiş tarayıcılarından daha iyi ve *daha çeşitli* bir yetenek havuzu oluşturabilir.

Bu örnekler ve diğerleri bizi kaçınılamaz bir çıkarıma götürüyor: Belirsiz bir dünyada kestirimci bir algoritmanın kusursuz olması çok olası değilken, gürültülü ve sıklıkla ön yargılı insan yargısından çok daha az kusurlu olması mümkündür. Bu üstünlük hem geçerlilik (iyi algoritmalar neredeyse her zaman daha iyi tahmin eder) hem de ayrımcılık (iyi algoritmalar insan yargıçlardan daha az ön yargılı olabilirler) açısından geçerlidir. Algoritmalar insan uzmanlardan daha az hata yapıyorsa ve bununla birlikte sezgisel tercihimiz insanlardan yana ise o zaman sezgisel tercihlerimiz dikkatle incelenmelidir.

Daha geniş çıkarımlarımız basit ve algoritmalar konusunun ötesine uzanıyor. Gürültüyü azaltma stratejilerinin maliyetli olabildiği doğru. Ama çoğu zaman bu maliyetler sadece bir bahaneden ibaret oluyor ve gürültünün adaletsizliğini ve maliyetlerini hoş görmek için yeterli neden oluşturmuyor. Elbette gürültüyü azaltma çabaları da belki de ön yargı biçiminde, kendilerine ait sorunlar üretebilir. Bu durumda ciddi bir sorunumuz var demektir ama çözüm gürültüyü azaltma çabalarını tümden bir kenara itmek değil, daha iyilerini bulmaktır.

Gürültüyü Azaltmanın Maliyetlerinden Bahsederken

"Eğitimde gürültüyü ortadan kaldırmaya çalışsaydık, çok para harcamamız gerekirdi. Öğretmenler öğrencilere not verirken gürültülüdür. Beş öğretmenin aynı kâğıda not vermesini sağlayamayız."

"İnsan yargısına bel bağlamak yerine, bir sosyal ağ hiç kimsenin belli sözcükleri hangi bağlamda olursa olsun kullanamayacağına karar verirse, bu, gürültüyü ortadan kaldırır ama aynı zamanda birçok hata yaratır. Tedavi hastalıktan beter olabilir."

"Ön yargılı kurallar ve ön yargılar olduğu doğru. Ama insanların da ön yargıları var. Asıl sormamız gereken, 'Hem gürültüsüz hem daha az ön yargılı algoritmalar tasarlayabilir miyiz?' olmalıdır."

"Gürültüyü ortadan kaldırmak maliyetli olabilir ama maliyet genellikle maruz kalmaya değer. Gürültü son derece adaletsiz olabilir. Ve eğer bir gürültüyü azaltma çabası çok hamsa –kendimizi kabul edilemeyecek kadar katı ya da kasıtsız olarak ön yargı üreten kılavuzlar ya da kurallarla karşı karşıya bulursak– pes etmemeliyiz. Tekrar denemeliyiz."

BÖLÜM 27

Saygınlık

Bir ev kredisi başvurunuzun, biri durumunuzu incelediği için değil, bankanın sizin kredi puanınıza sahip insanların kredi alamayacağı yönünde katı bir kuralı olduğu için reddedildiğini varsayın. Ya da harika özellikleriniz var ve bir şirketteki bir mülakatçı sizden çok etkilendi ama iş başvurunuz on beş sene önce bir uyuşturucu suçundan hüküm giydiğiniz ve şirketin herhangi bir hüküm giymiş hiç kimseyi işe almamak gibi bir kuralı olduğu için reddedildi. Ya da belki de bir şeyle suçlanıyorsunuz ve kefalet talebiniz gerçek bir insanın karşısında size özel bir oturumun sonunda değil, bir algoritma sizin özelliklerinize sahip insanların kefalete izin veren eşiği aşan bir firar riskinin olduğuna karar verdiği için reddedildi.

Bu durumlarda, pek çok insan itiraz eder. Birey olarak muamele görmek isterler. Gerçek bir insanın onlara özel duruma bakmasını isterler. Bireyselleştirilmiş muamelenin gürültü üreteceğinin farkında olabilirler de olmayabilirler de. Ama bu tür bir muamelenin bedeli buysa, o bedeli ödemeye değeceğinde ısrar ederler. İnsanların itiraz ettiği, Yüksek Mahkeme'nin sözleriyle, "benzersiz bireyler olarak değil", herhangi bir cezaya "körü körüne

çarptırılacak, sureti olmayan, birbirinden ayırt edilemez bir kütle olarak muamele etmesidir." (bakınız Bölüm 26).

Pek çok kişi, insanlara birey olarak ve dolayısıyla bir tür saygıyla muamele gördükleri hissini vermek için, kuralların zorbalığından arındırılmış olarak gördükleri bireysel oturumlarda ısrarcıdır. Sıradan bir hayatın parçası olarak görülen yargı süreci, takdir yetkisini kullanma yetkisine sahip bir insanın çok geniş bir faktör yelpazesini değerlendirdiği yüz yüze bir etkileşim fırsatını gerektirir gibi görünebilir.

Pek çok kültürde, vakaya göre yargı argümanının derin ahlaki temelleri vardır. Politikada, hukukta, teolojide ve hatta edebiyatta bulunabilir. Shakespeare'in *Venedik Taciri* gürültüsüz kurallara bir itiraz ve hukukta ve genel olarak insan yargısında merhametin rolü için bir yakarı olarak yorumlanabilir. Portia'nın kapanış argümanı şöyledir:

Merhamet zorla olmaz;
Gökten süzülen yağmur gibidir.
İki yönden kutsaldır:
Hem vereni kutsal kılar hem alanı.
(...)
Kralların yüreğinde taht kurmuştur o;
Yüce Tanrı'nın bir özelliğidir.
Dünyasal gücün Tanrısal güce en yakın hâli.
Adaletle merhametin uzlaşmasıyla ortaya çıkar.[1]

Merhamet kurallara bağlı olmadığı için gürültülüdür. Buna rağmen, Portia'nın yakarısı pek çok durumda ve sayısız örgütte yapılabilir. Sık sık yankı bulur. Bir çalışan terfi almaya çalışıyor olabilir. Ev sahibi olmak isteyen bir kişi krediye başvuruyor ola-

1 *Venedik Taciri*, çev: Bülent Bozkurt, Remzi Kitabevi, 1992. (yay. n.)

bilir. Bir öğrenci üniversiteye başvuruyor olabilir. Bu tür vakalar için karar alan kişiler, bazı gürültü azaltıcı stratejileri bütün katı kuralların üstünde reddedebilir. Reddetmezlerse, bunun nedeni Portia'nın dediği gibi, merhametin zorla olmadığını düşünmeleri olabilir. Yaklaşımlarının gürültülü olduğunu bilebilirler ama o yaklaşım insanlara saygıyla muamele gördüklerini ve birinin onları dinlediğini hissettirecekse, onu yine de kabullenebilirler.

Bazı gürültüyü azaltma stratejileri bu itiraza rastlamaz. Bir kararı sadece bir değil, üç kişi birlikte veriyorsa, insanlara yine de bireyselleştirilmiş bir oturum hakkı sağlanıyordur. Kılavuzlar karar alıcılara önemli takdir yetkileri bırakabilir. Ancak katı kurallar dâhil birtakım gürültüyü azaltma çabaları, o takdir yetkisini ortadan kaldırır ve insanların bunun sonucu olan sürecin saygınlık hislerine hakaret olduğunu söyleyerek itiraz etmelerine yol açabilir.

Haklı mıdırlar? Elbette insanlar bireyselleştirilmiş bir oturum fırsatı bulmayı genellikle önemserler. Dinlenme fırsatının tartışılmaz bir insani değeri vardır. Ancak eğer bireyselleştirilmiş oturumlar daha fazla ölüme, daha fazla haksızlığa ve daha yüksek maliyetlere neden oluyorsa, yüceltilmemeliler. İşe alım, okula kabul ve tıp gibi durumlarda bazı gürültüyü azaltma stratejileri ham olabilir, gürültülü olmakla birlikte dengede daha az hata üreten bireyselleştirilmiş muamele türlerini menedebilirler. Ama eğer bir gürültüyü azaltma stratejisi hamsa, o zaman daha önce de belirttiğimiz gibi, en iyi tepki daha iyi bir strateji, çok geniş bir alakalı değişkenler aralığına uyacak bir strateji üretmek olacaktır. Ve eğer o daha iyi strateji gürültüyü ortadan kaldırır ve daha az hata üretirse, dinlenme fırsatını ortadan kaldırsa bile, bireyselleştirilmiş muameleye bariz üstünlükleri olacaktır.

Bireyselleştirilmiş muameleye ilginin önemli olmadığını söylemiyoruz. Ama eğer bu tür bir muamele, elle tutulur haksızlığın da aralarında olduğu her tür korkunç sonuca yol açıyorsa, yüksek bir bedeli olacaktır.

Değerleri Değiştirmek

Bir kamu kuruluşunun gürültüyü ortadan kaldırmayı başardığını düşünün. Bir üniversite, *görevi kötüye kullanma*yı her öğretim görevlisi ve her öğrencinin neyi içerip neyi içermediğini bileceği şekilde tanımlamış olsun. Ya da büyük bir şirket, şirkette çalışan herkesin neye izin verildiğini, neyin yasak olduğunu açıkça bilmesi için, *yolsuzluğun* ne anlama geldiğini kesin olarak tanımlasın. Ya da özel bir kurumun, üniversitede belirli alanlardan diploma almamış kişileri işe almayacağını ilan ederek gürültüyü önemli ölçüde azalttığını hayal edin. Ya örgütün değerleri değişirse, o zaman ne olur? Bazı gürültüyü azaltma stratejileri onlara yer açamayabilir ve katılıkları, bireyselleştirilmiş muamele ve saygınlığa ilgiyle yakından bağlantılı bir sorun yaratabilir.

Amerikan anayasa hukukunda alınmış çok ünlü ve şaşırtıcı bir karar bu noktayı anlamamızı kolaylaştırıyor. 1974'te alınan karara konu olan dava, bir okulun hamile öğretmenlerin bebeğin beklenen doğum tarihinden beş ay önce ücretsiz izne çıkmasını talep eden katı kuralıyla ilgiliydi. Jo Carol LaFleur adında bir öğretmen, öğretmenlik yapmaya son derece uygun olduğunu, kuralın ayrımcı olduğunu ve beş ayın abartılı olduğunu savunuyordu.

ABD Yüksek Mahkemesi de aynı görüşteydi. Ama mahkeme cinsiyet ayrımcılığından ya da beş ayın illa aşırı olduğundan bahsetmedi. Bunların yerine, LaFleur'e, özellikle *onun* çalışmayı bırakması için fiziksel bir ihtiyaç olmadığını gösterme fırsatının verilmediğini söyleyerek itiraz etti. Mahkemenin kendi ifadesiyle,

> *Öğretmenin doktorunun –ya da okul yönetim kurulunun– belli bir öğretmenin işine devam etme becerisi konusunda bireyselleşmiş bir saptaması bulunmuyor. Kurallar çürütülemez bir fiziksel yetersizlik varsayımı içeriyor ve bu varsayım belli bir kadının fiziksel durumuyla ilgili tıbbi kanıtlar tamamen aksini gösterirken bile geçerliliğini koruyor.*

Beş aylık zorunlu bir süreç gerçekten saçma görünüyor. Ama mahkeme bu noktayı vurgulamadı. Onun yerine, "çürütülemez varsayım"dan ve "bireyselleşmiş saptama"nın yokluğundan yakındı. Mahkeme bunu söyleyerek, açıkça, Portia gibi, merhametin zorla olmadığını ve LaFleur'ün kendine özel şartlarının gözden geçirilmesinin gerektiğini savunmuş oldu.

Ancak bir karar hijyeni olmadığı sürece, bu bir gürültü reçetesinden başka bir şey değildir. LaFleur'ün vakasında kim karar verecek? Onun için alınan karar, benzer durumdaki daha pek çok kadın için aynı olacak mı? Her durumda, birçok kural çürütülemez varsayımlara neden olur. Belirlenen bir hız sınırı kabul edilebilir midir? Oy kullanma ya da içki içme için asgari bir yaş? Alkollüyken araç kullanmaya kesin bir yasak? Eleştirmenler bu tür örneklere dayanarak "çürütülemez varsayımlar"a karşı bir argümanın, özellikle amaçları ve etkileri gürültüyü azaltmak olduğu için, çok şey ispatlayacağını söyleyerek itiraz ettiler.

O dönemin nüfuzlu yorumcuları ahlaki değerlerin zamanla değiştiğini ve dolayısıyla katı kurallardan kaçınılması gerektiğini vurgulayarak mahkemenin kararını savundular. Kadınların toplumdaki rolü açısından, sosyal normların büyük bir değişim hâlinde olduğunu belirttiler. Bireyselleştirilmiş saptamaların bu bağlamda özellikle uygun olduğunu, çünkü değişen normların da dâhil edilmesine olanak sağlayacaklarını vurguladılar. Kurallara bağlı bir sistem gürültüyü ortadan kaldırabilirdi, bu iyi bir şeydi ama aynı zamanda mevcut norm ve değerleri dondurabilirdi ve bu o kadar iyi olmazdı.

Kısaca, bazı insanlar gürültülü bir sistemin avantajının insanların yeni ve yeni yeni ortaya çıkan değerleri de uyarlamasına izin vermesi olduğunda ısrar edebilirler. Değerler değiştikçe, eğer yargıçların takdir yetkisi kullanmalarına izin verilirse, örneğin, uyuşturucu suçlarından hüküm giyenlere daha düşük, tecavüzden hüküm giyenlere daha yüksek cezalar vermeye başlayabilirler. Bazı

yargıçlar daha yumuşaksa ve diğerleri değilse ortada bir derece haksızlık olacağını vurgulamıştık; benzer durumdaki insanlar farklı muamele görürdü. Ancak yeni ve yeni yeni ortaya çıkan sosyal değerlere yer verdiği sürece haksızlık hoş görülebilir.

Sorun ceza yargı sistemi ya da hukukun geneliyle sınırlı değil. Şirketler herhangi bir politika için, gürültü üretecek bile olsa yargı ve kararlarında esnekliğe izin vermeye karar verebilirler çünkü esneklik yeni inançlar ve değerler doğdukça politikaların zaman içinde değiştirilebilmesini sağlar. Kişisel bir örnek verelim: Aramızdan biri birkaç yıl önce büyük bir danışmanlık şirketine katıldığında, şirketten geri ödeme talep etme hakkının olduğu seyahat masraflarının çok eski tarihli olmayan bir hoş geldin paketiyle belirlendiğini öğrendi ("gideceği yere sapasağlam vardığını haber vermek için eve açacağı bir telefon, takım elbisesinin ütü masrafı, valizlerini taşıyan otel görevlisine bahşiş"). Kurallar gürültüsüzdü ama açıkça modası geçmişti (ve cinsiyetçi). Çok geçmeden yerlerine zamanla geliştirilebilecek standartlar geldi. Örneğin, masraflar artık "gerekli ve makul" olmalıydı.

Gürültünün bu savunmasına ilk cevap çok basit: Kimi gürültüyü azaltma stratejileri bu itirazla hiç karşılaşmaz. Eğer insanlar bir dış bakışa dayalı ortak bir ölçek kullanırlarsa, zaman içinde değişen değerlere tepki verebilirler. Her durumda, gürültüyü azaltma çabalarının kalıcı olmasına gerek yoktur ve olmamalıdır da. Bu tür çabalar katı kurallar biçimini alırsa, o kuralları koyanlar zaman içinde değişiklik yapmaya istekli olmalıdır. O kuralları yıllık olarak gözden geçirebilirler. Yeni değerler nedeniyle yeni kuralların gerekli olduğuna karar verebilirler. Ceza yargı sisteminde kural koyucular belirli suçlar için hüküm sürelerini azaltırken, diğerlerini artırabilirler. Bazı aktiviteleri suç olmaktan tamamen çıkarabilir ve daha önce tamamen kabul edilebilir görülen bir aktiviteyi suç olarak belirleyebilirler.

Ama bir adım geri çekilelim. Gürültülü sistemler yeni yeni ortaya çıkan ahlaki değerlere yer açabilir ve bu iyi bir şey olabilir. Ancak pek çok alanda yüksek gürültü seviyelerini bu argümanla savunmak abes olacaktır. Yargıları bir araya toplamak gibi, en önemli gürültüyü azaltma stratejilerinin bazıları yeni yeni ortaya çıkan değerlere yerlerini bırakmaktadır. Ve eğer arızalanan bir dizüstü bilgisayardan yakınan farklı müşteriler bir bilgisayar firmasından farklı muameleler görürse, tutarsızlığın yeni yeni ortaya çıkan değerlerden kaynaklanması pek olası değildir. Farklı insanlara farklı tıbbi teşhisler konması, nadiren yeni ahlaki değerlerle alakalıdır. Gürültüyü azaltmak ve hatta tamamen ortadan kaldırmak için çok şey yaparken, bir taraftan da değerlerin gelişmesine olanak sağlayacak süreçler tasarlayabiliriz.

Sistemi Manipüle Etmek, Kurallardan Kaçmak

Gürültülü bir sistemde, her tür yargı durumun gerektirdiği gibi uyarlanabilir ve beklenmedik gelişmelere cevap verebilir. Bazı gürültüyü azaltma stratejileri, adaptasyonun gücünü ortadan kaldırarak, istenmeyecek bir şekilde, insanları sistemi manipüle etmeye özendirme sonucuna yol açabilirler. Gürültüyü hoş görmenin potansiyel bir savunusu, özel ve kamusal kurumların bu tür bir manipülasyonu önlemek için benimsediği yaklaşımların bir yan ürünü olabilmesidir.

Vergi kanunu bunun tanıdık bir örneğidir. Bir yandan, vergi sistemi gürültülü olmamalıdır. Net ve öngörülebilir olmalı, aynı durumdaki veri mükelleflerine farklı muamele edilmemelidir. Ama vergi sisteminde gürültüyü ortadan kaldırsaydık, akıllı vergi mükellefleri kaçınılmaz olarak kurallardan kaçmanın bir yolunu bulurlardı. Vergi uzmanları arasında, gürültüyü ortadan kaldıran net kurallar koymanın mı, yoksa öngörülmezliğe izin veren ama aynı zamanda açık kuralların, fırsatçı ya da kişinin çıkarını kollayan

davranışlar üretme riskini azaltan belli bir derece müphemliğe yer bırakmanın mı daha iyi olacağı tartışması sürmektedir.

Bazı şirket ve üniversiteler insanları "haksızlık" yapmaktan, ne anlama geldiğini tam olarak tanımlamadan menederler. Bunun kaçınılmaz sonucu iyi olmayan ve hatta çok kötü olabilecek gürültüdür. Ancak nelerin "kabahat" sayıldığını belirleyen kesin bir liste olsa, bu kez, listenin açıkça kapsamadığı korkunç davranışların hoş görülmesi mümkün olacaktır.

Kuralların kesin çizgileri olduğu için, insanlar teknik olarak hariç tutulan ama aynı ya da benzer zararlara yol açan davranışlara girerek, kurallardan kaçabilirler. (Bunu bütün ergen ebeveynleri bilir!) Yasaklanması gereken bütün davranışları menedebilecek kuralları kolayca tasarlayamadığımız zaman, gürültüyü hoş görmek için belirgin bir nedenimiz olur ya da en azından itiraz böyle söyler.

Bazı durumlarda, gürültüyü ortadan kaldıran açık ve tanımlanmış kurallar kaçınma riskine sebep olurlar. Ve bu, gürültüyü azaltmak için bir araya toplama gibi başka bir strateji benimsemek ve belki de bir miktar gürültüye izin veren bir yaklaşımı hoş görmek için bir neden olabilir. Ama sözcükler hayati önem *taşıyabilir.* Ne kadar kaçınma olacağını ve ne kadar gürültü olacağını sormalıyız. Sadece çok az kaçınma ve çok gürültü varsa, o zaman gürültüyü azaltan yaklaşımlarla daha iyi yoldayız demektir. Bu soruya Bölüm 28'de geri döneceğiz.

Caydırıcılık ve Riskten Kaçınma

Amacın çalışanlar, öğrenciler ya da sıradan vatandaşlar için suistimalden caydırmak olduğunu varsayalım. Biraz öngörülemezlik, hatta belki de bolca öngörülemezlik en kötü şey olmayabilir. Bir işveren, "Belli kabahatlerin cezalandırılması para cezası, uzaklaştırma ya da kovulmak olursa çalışanlarım bu tür bir kabahat işlemeyecektir", diye düşünebilir. Bir ceza yargı sistemini yöne-

tenler, "Farazi suçluların olası cezaları tahmin etmek zorunda kalmasını çok önemsemiyoruz. Eğer bir ceza piyangosu ihtimali, kişileri çizgiyi aşmaktan caydırıyorsa o zaman belki de ortaya çıkan gürültü hoş görülebilir", diye düşünebilirler.

Soyut olarak, bu argümanlar geçiştirilemezler ama korkunç ikna edici de değildirler. İlk bakışta önemli olan cezanın beklenen değeridir ve 5.000 dolarlık bir cezanın %50 olasılığı, 2.500 dolarlık bir cezanın kesinliğine denktir. Elbette bazı insanlar en kötü durum senaryosuna odaklanabilirler. Riskten kaçınan insanlar için 5.000 dolarlık bir cezanın %50 ihtimali daha caydırıcı olabilir ama riske açık insanlar için o kadar caydırıcı olmayacaktır. Gürültülü bir sistemin daha fazla caydırıcılık sağlayıp sağlamadığını bilmek için, potansiyel kabahatlilerin riskten kaçınan ya da riske açık insanlar olduğunu bilmemiz gerekir. Ve caydırıcılığı artırmak istiyorsak, cezayı artırıp gürültüyü ortadan kaldırmak daha iyi olmaz mı? Bunu yapmak haksızlığı da ortadan kaldıracaktır.

Yaratıcılık, Moral ve Taze Fikirler

Bazı gürültüyü azaltma çabaları motivasyonu ve bağlılığı ezebilir mi? Yaratıcılığı etkileyip insanların büyük başarılar yakalamasını önleyebilirler mi? Pek çok örgüt böyle düşünüyor. Bazı örneklerde haklı olabilirler. Haklı olup olmadıklarını anlamak için, itiraz ettikleri gürültüyü azaltma stratejisini belirlemeliyiz.

Birçok yargıcın cezai hüküm kılavuzlarına yoğun olumsuz tepkisini hatırlayın. Bir yargıcın ifadesiyle, "Mahkeme salonunda yargının uygulanmasına güvenmeyi yeniden öğrenmeliyiz." Genel olarak, otorite sahibi insanlar takdir yetkilerinin ellerinden alınmasından hoşlanmazlar. Kendilerini hem eksiltilmiş hem de kısıtlanmış, hatta küçük düşürülmüş hissedebilirler. Takdir yetkilerini azaltmak için adımlar atıldığı zaman, pek çok insan isyan eder. Yargı kullanma fırsatına değer verirler, hatta el üstünde tutuyor

da olabilirler. Diğer herkesin yaptıklarını yapmaları için takdir yetkileri ellerinden alınırsa, kendilerini bir makinenin dişlileri gibi hissedebilirler.

Kısaca, gürültülü bir sistem, gürültülü olduğu için değil, insanların uygun gördükleri şekilde karar vermelerine imkân tanıdığı için moral açısından iyi gelebilir. Çalışanların müşteri şikâyetlerini kendilerine göre cevaplamalarına, altlarında çalışan kişileri gönüllerince değerlendirmelerine ya da primleri uygun gördükleri şekilde belirlemelerine izin verilirse, işlerinden daha fazla keyif alabilirler. Şirket gürültüyü azaltmak için adımlar atarsa, çalışanlar kendi yetkilerinden ödün verildiğini düşünebilirler. Artık, yaratıcılıklarını uygulamıyor sadece kurallara uyuyor olurlar. İşleri daha mekanik ve robotik görünür. Kim bağımsız kararlar alma kapasitesini çiğneyen bir yerde çalışmak ister ki?

Örgütler bu duygulara sadece onları onurlandırmak için değil, insanlara yeni fikirler üretecek alan vermek istedikleri için de karşılık verebilirler. Bir kuralın konması hüner ve yeni keşifleri azaltabilir.

Bu noktalar pek çok örgüte uyabilir ama elbette hepsine uymaz. Farklı görevler farklı şekillerde değerlendirilmelidir; gürültülü boğaz enfeksiyonu ya da yüksek tansiyon teşhisleri yaratıcılık kullanmak için iyi bir yer olmayabilir. Ama daha mutlu ve ilham verici bir iş gücüne imkân veriyorsa, gürültüyü hoş görmeye razı olabiliriz. Moralsizlik başlı başına bir maliyettir ve kötü performans gibi, başka maliyetler doğurur. Bir yandan gürültüyü azaltırken, bir yandan taze fikirlere açık kalabilmeliyiz. Karmaşık yargıları yapılandırmak gibi, bazı gürültüyü azaltma stratejileri, tam olarak bunu yaparlar. Morali yüksek tutarken gürültüyü azaltmak istiyorsak, bu sonucu sağlayan karar hijyeni stratejilerini seçebiliriz. Ve sorumlu kişiler, katı kurallar yürürlükte olduğunda bile, onlara meydan okuyacak ve üstlerinde düşünmeye imkân verecek bir

sürecin var olduğunu ama vakaya göre takdir yetkisi kullanarak kuralları çiğnemenin mümkün olmadığını açıkça belirtebilirler.

Saygın bir hukukçu ve düşünür olan Philip Howard bir dizi enerjik kitapta, daha esnek yargılara izin vermenin lehine benzer noktalara dikkat çekmiştir. Howard politikaların gürültüyü azaltan buyurgan kurallarla değil, "akılcı olmak", "tedbirli davranmak" ve "aşırı riskler dayatmamak" gibi genel prensiplerle şekillenmesini ister.

Howard'a göre, hükûmet düzenlemelerinin modern dünyası, sadece çok katı olması nedeniyle, çıldırmıştır. Öğrenciler, çiftçiler, geliştiriciler, hemşireler, doktorlar, bütün bu uzmanlar ve çok daha fazlası onlara ne yapacaklarını ve tam olarak nasıl yapacaklarını söyleyen yüklerin ağırlığı altında ezilmektedirler. Howard söz konusu hedef ister eğitimde daha iyi sonuçlar, daha az kaza, daha temiz su ya da daha sağlıklı hastalar olsun, insanların, ilgili hedeflere ulaşmanın yolunu bulmak için kendi yaratıcılıklarını kullanmalarına izin vermenin daha iyi olacağını düşünür.

Howard birtakım cazip argümanlar öne sürmektedir ama savunduğu yaklaşımların gürültü ve ön yargıda potansiyel artışları da içeren sonuçlarını sorgulamak önemlidir. Çoğu insan soyut bir kavram olarak katılıktan hoşlanmaz ama gürültüyü azaltmanın, ön yargı ve hatayı bertaraf etmenin en iyi yolu bu olabilir. Sadece genel prensipler yürürlükte olursa, bunu, yorumlanmalarında ve uygulanmalarında gürültü takip edecektir. O gürültü katlanılmaz hatta skandal düzeyinde olabilir. En azından, gürültünün maliyetlerine dikkatle kafa yorulmalıdır; ki genellikle yorulmaz. Gürültünün yaygın haksızlığa ve kendi başına yüksek maliyetlere neden olduğunu gördüğümüz zaman, genellikle kabul edilemez olduğuna ve önemli değerlerden ödün vermeyen gürültüyü azaltma stratejileri belirlememiz gerektiğine karar veririz.

Saygınlıktan Bahsederken

"İnsanlar yüz yüze etkileşimlere değer verir, hatta ihtiyaç duyarlar. Gerçek bir insanın endişelerini ve şikâyetlerini dinlemesini ve bir şeyleri düzeltme gücüne sahip olmasını isterler. Elbette bu etkileşimler, kaçınılmaz olarak gürültü üretecektir. Ama insan saygınlığı paha biçilmezdir."

"Ahlaki değerler sürekli olarak değişir ve gelişir. Her şeye kilit vurursak değişen değerlere yer açamayız. Bazı gürültüyü azaltma çabaları çok katı oldukları için ahlaki değişimi önleyeceklerdir."

"İnsanları suistimalden caydırmak istiyorsanız, bir miktar gürültüyü hoş görmelisiniz. Eğer öğrenciler intihalin cezası konusunda net değillerse, harika, bu onları intihalden uzak tutacaktır. Gürültü biçiminde bir miktar belirsizlik, caydırıcılığı artırabilir."

"Gürültüyü ortadan kaldırırsak, karşımızda yanlışı tercih edenlerin kaçınmanın bir yolunu illa bulacağı kesin kuralları bulabiliriz. Gürültü, stratejik ve fırsatçı davranışı önlemenin bir yoluysa, ödenmeye değer bir bedel olabilir."

"Yaratıcı insanların alana ihtiyacı vardır. İnsanlar robot değildir. İşiniz ne olursa olsun, manevra yapacak alana ihtiyaç duyarsınız. Etrafınız kuşatılırsa, gürültülü olmayabilirsiniz ama çok eğlenemez ve orijinal fikirler üretemeyebilirsiniz."

"Sonunda, gürültüyü savunan çabaların çoğu ikna edici değil. İnsanların saygınlığına saygı duyabilir, ahlaki değişim ve gelişime yer bırakabilir ve haksızlığı ve gürültünün maliyetini hoş görmeden, insan yaratıcılığına fırsat sağlayabiliriz."

BÖLÜM 28

Kurallar mı Standartlar mı?

Amaç gürültüyü azaltmak ya da bunun nasıl yapılacağına veya yapılıp yapılmayacağına (ve hangi dereceye kadar) karar vermekse iki düzenleyici davranış arasındaki farkı ayırt etmek faydalı olacaktır: Kurallar ve standartlar. Her türden örgüt bunlardan birini ya da ikisinin bir kombinasyonunu seçerler.

İş dünyasında, şirketler çalışanlarının belirli saatler arasında iş yerinde olmak zorunda olduğunu, kimsenin iki haftadan daha uzun süre izne çıkamayacağını ve basına bilgi sızdıran kişinin işten kovulacağını söyleyebilir. Alternatif olarak, çalışanların "mantıklı bir çalışma saati süresince" iş yerinde olmasını, izinlere "duruma göre ve şirketin ihtiyaçlarıyla tutarlı olarak ayrı ayrı" karar verileceğini ve basına bilgi sızdıran kişinin "uygun şekilde cezalandırılacağını" da söyleyebilir.

Hukukta bir kural hiç kimsenin sayısal bir hız sınırını aşamayacağını, işçilerin kanserojen maddelere maruz bırakılamayacağını ya da reçeteli bütün ilaçların beraberinde belirli uyarılar olması gerektiğini söyleyebilir. Tam tersine, bir standart insanların araçlarını "temkinli" kullanmasını, çalışanlara "uygulanabilirlik ölçüsünde" güvenli iş yerleri sağlanması gerektiğini ya da reçeteli

ilaçlar konusunda uyarı sağlamada şirketlerin "makul" hareket etmesi gerektiğini söyleyebilir.

Bu örnekler bize kurallar ve standartlar arasındaki temel farkı göstermektedir. Kurallar onları uygulayanların takdir yetkisini ortadan kaldırmayı amaçlar; standartlar ise bu takdir yetkisini bağışlamayı. Kurallar yürürlükte olduğu zaman, gürültünün ciddi anlamda azalması beklenir. Kuralları yorumlayanlar bir olgu sorusunu cevaplamalıdır: Sürücü ne kadar hızlı gidiyordu? Bir işçi kanserojen maddeye maruz kaldı mı? İlaç gerekli uyarıları taşıyor muydu?

Kurallar altında, olgu bulma girişiminin kendisi yargı içerebilir ve dolayısıyla gürültü üretebilir ya da ön yargıdan etkilenebilir. Birçok örneğe rastladık. Ancak kuralları tasarlayan insanlar bu riskleri azaltmayı amaç edinirler ve bir kural bir sayıyı içerdiği zaman ("hiç kimse on sekiz yaşını doldurmadan oy kullanamaz" ya da "hız sınırı saatte 100 km'dir") gürültünün azalması beklenir. Kuralların önemli bir özelliği vardır: *Yargının rolünü azaltırlar.* En azından bu açıdan, yargıçlara (kuralları uygulayan herkesi içerecek şekilde anlaşıldığında) daha az iş düşer. Kurallara uyarlar. Konu ne olursa olsun, manevra yapacak çok daha az alanları olur.

Standartlar ise tamamen farklıdır. Standartlar yürürlükte olduğunda, açık uçlu terimlerin anlamını belirlemek konusunda yargıçlara çok iş düşer. Örneğin neyin "makul" ve "yapılabilir" sayıldığına karar vermek için sayısız yargıda bulunmaları gerekir. Olguları bulmaya ek olarak, nispeten muğlak ifadelere içerik katmaları icap eder. Standartları tasarlayanlar karar alma yetkisini etkili bir şekilde başkalarına verirler. Gücü delege ederler.

Bölüm 22'de değindiğimiz kılavuz türleri kurallar ve standartlar olabilir. Eğer kural iseler, yargıyı çarpıcı bir şekilde sınırlandırırlar. Standart bile olsalar açık uçlu olmaya uzak olabilirler. Apgar puanlamaları kılavuzdur ve kural değildir. Takdir yetkisinin kullanılmasını yasaklamazlar. Kılavuzlar takdir yetkisini ortadan

kaldıracak kadar sıkıldıkları zaman kurallara dönerler. Algoritmalar standart olarak değil, kural olarak çalışır.

Bölmeler ve Bilgisizlik

Şirketlerin, örgütlerin, toplulukların ya da grupların keskin bir şekilde bölündüğü durumlarda, standart üretmenin kural üretmekten çok daha kolay olduğu en başından net olmalıdır. Şirket liderleri, yasağın tam olarak ne anlama geldiğini bilmeden, yöneticilerin taciz edici bir şekilde davranmaması gerektiği konusunda hemfikir olabilirler. Yöneticiler flörtçü davranışların kabul edilebilir olup olmadığına karar vermeden, iş yerinde cinsel tacize karşı çıkabilirler. Bir üniversite, bu terimin tam anlamını belirtmeden öğrencilerinin intihale yeltenmesini yasaklayabilir. İnsanlar ticari reklamları, tehditleri ya da müstehcenliği koruyup korumayacağına karar vermeden, bir anayasanın ifade özgürlüğünü korumak gerektiği konusunda anlaşmaya varabilir. İnsanlar ihtiyatın ne anlama geldiğini tanımlamadan, çevre düzenlemelerinin sera gazı emisyonlarını azaltacak ihtiyatlı kurallar içermesini kabul edebilirler.

Detay vermeden standart belirlemek gürültüye yol açabilir ve bu gürültü, yargıları bir araya toplamak ya da aracı değerlendirme protokolü kullanmak gibi, ele aldığımız stratejilerin bazıları kullanılarak kontrol edilebilir. Liderler kurallar koymak isteyebilirler ama pratikte görüş birliği sağlayamayabilirler. Anayasalar pek çok standart içerir (örneğin din özgürlüğünü koruyan standartlar). Aynı durum İnsan Hakları Evrensel Bildirisi için de geçerlidir. ("Bütün insanlar özgür, onur ve haklar bakımından eşit doğarlar.")

Çeşitli insanların gürültüyü azaltıcı kurallar üzerinde anlaşmalarını sağlamanın büyük zorluğu, kuralların değil standartların yürürlüğe konmasının bir nedenidir. Bir şirketin liderleri çalışanların müşterilerle nasıl ilgileneceğini kontrol etmek için belirli sözcükler konusunda anlaşmayı başaramayabilirler. Standartlar

bu liderlerin elinden gelenin en iyisi olabilir. Kamu sektöründe de benzer durumlar mevcuttur. Bir kanunu yürürlüğe koymanın bedeli buysa, kanun yapıcılar bir standart üstünde uzlaşmaya varabilirler (ve doğacak gürültüyü hoş görebilirler). Tıpta, doktorlar hastalıkları teşhis etmek için standartlar üstünde anlaşabilir; öte yandan kurallar geliştirmek inatçı anlaşmazlıklara yol açabilir.

Ancak sosyal ve politik bölünmeler insanların kurallar yerine standartlara sığınmasının tek nedeni değildir. Bazen asıl sorun insanların sağduyulu kararlar üretmelerine olanak sağlayacak bilgiye sahip olmamaları olabilir. Bir üniversite, öğretim görevlilerini terfi ettirme kararlarını yönetecek kurallar üretmeyi beceremeyebilir. Bir işveren çalışanları elinde tutmasına ya da disipline etmesine yol açacak bütün şartları öngörmekte zorlanabilir. Ulusal bir yasama organı havayı kirleten maddelerin –parçacıklı madde, ozon, nitrojen dioksit, kurşun– kabul edilebilir seviyesi hakkında bilgi sahibi olmayabilir. Yapılabilecek en iyi şey bir standart yayımlamak ve sonucu gürültü bile olsa, anlamını netleştirmek konusunda güvenilir uzmanlara bel bağlamaktır.

Kurallar birçok şekilde ön yargılı olabilir. Bir kural kadınları polis memuru olmaktan menedebilir. İrlandalıların başvurmamasını söyleyebilir. Büyük ön yargılar yaratsalar da, kurallar (herkes uyarsa) gürültüyü keskin bir şekilde azaltır. Eğer bir kural yirmi bir yaş üstü herkesin alkollü içki almasına izin olduğunu ve yirmi bir yaş altı kimsenin alamayacağını söylüyorsa, muhtemelen az gürültü olacaktır; en azından insanlar kurala uyduğu sürece. Aksine, standartlar gürültüyü davet eder.

Patronlar, Astları Kontrol Etmek

Kurallar ve standartlar arasındaki ayrımın bütün kamusal ve her tür iş alanının da dâhil olduğu özel kurumlarda önemi çok büyüktür. Ne zaman bir sorumlu uygulayıcı bir personeli kontrol

etmek istese, bu ikisi arasında seçim yapma ihtiyacı doğar. Bölüm 2'de anlatıldığı gibi, risk değerlendirme uzmanları şirketlerine fayda sağlamak için en makul primi (ne çok yüksek ne çok düşük) teklif etmek için çok çalışırlar. Patronları bu uzmanlara rehberlik etmesi için standartlar mı, yoksa kurallar mı sunar? Bir şirketteki herhangi bir lider çalışanları çok belirgin ya da daha genel bir şekilde ("sağduyunuzu kullanın" ya da "en iyi yargınızı ortaya koyun") yönlendirebilir. Bir doktor hastasına talimat verirken bu iki yaklaşımdan birini kullanabilir. "Her sabah ve her akşam bir hap alın." bir kuraldır. "İhtiyacınız olduğunu hissettiğiniz zaman bir hap alın." ise bir standart.

Facebook gibi sosyal medya şirketlerinin kaçınılmaz olarak gürültüyü ve nasıl azaltılacağını dert edeceğinden bahsetmiştik. Şirket çalışanlarına net bir kuralı (örneğin, çıplaklığı yasaklayan) çiğneyen bir paylaşımı kaldırmalarını söyleyebilir. Ya da çalışanlarından bir standardı uygulamalarını da (zorbalığı ya da açıkça saldırgan materyalleri yasaklamak gibi) isteyebilir. Facebook'un ilk kez 2018'de ilan edilen Topluluk Standartları, kural ve standartların her ikisinden bolca bulunan büyüleyici bir karışımdır. Bu standartların duyurulmasının ardından, şirketin standartlarının aşırı miktarda gürültü ürettiğini (dolayısıyla hem hata hem haksızlık ürettiğini) savunan Facebook kullanıcılarından sayısız şikâyet geldi. Endişe, binlerce Facebook gözlemcisinin karar alması gerekeceği için kararların son derece değişken olacağıydı. Gözlemciler gözden geçirdikleri paylaşımları kaldırıp kaldırmamak konusunda karar verirken, neye izin olduğu, neyin yasak olduğu konusunda farklı yönde kararlar bildiriyorlardı. Böyle bir değişkenliğin neden kaçınılmaz olduğunu anlamak için, Facebook'un Topluluk Standartları'ndan 2020 yılına ait şu sözcükleri değerlendirin.

> *Nefret söylemini insanlara koruma altındaki özellikler olarak adlandırdığımız ırk, etnik köken, millî köken, en-*

gellilik durumu, dinî inanç, sınıf, cinsel yönelim, cinsiyet, cinsel kimlik ve ciddi bir hastalık gibi özelliklerinden dolayı doğrudan saldırıda bulunulması olarak tanımlıyoruz. Ayrıca göçmenlik statüsü için de korumalar sağlıyoruz. Saldırıları şiddet veya itibarsızlaştırıcı söylem, aşağılama ifadeleri, dışlama ya da ayrımcılık çağrıları olarak tanımlıyoruz.

Bu tür bir tanımı uygularken, gözlemcilerin gürültülü olması kaçınılmazdır. Tam olarak neler "Şiddet veya itibarsızlaştırıcı söylem" sayılır? Facebook bu tür soruların farkındaydı ve bunlara cevap olarak, tam olarak gürültüyü azaltmak için katı kurallara yöneldi. Bu kurallar kamuoyuna açık olmayan ve *New Yorker*'ın ele geçirdiği yaklaşık on iki bin kelimelik, Uygulama Standartları adlı bir belgede sınıflandırıldı. Halka açık Topluluk Standartları'nda, şiddet içeriğiyle ilk metin standart bir "Şiddeti yücelten içeriği kaldırıyoruz." cümlesini içeriyor. (Bu, tam olarak nedir?) Tezat olarak, Uygulama Standartları şiddet içeren görüntüleri listeledi ve içerik moderatörlerine bu görüntüler hakkında ne yapacaklarını açıkça bildirdi. Örnekler "yanmış veya yanan insan" ve "kendini yenileyemeyen vücut parçalarının kopması" gibi tanımları içeriyordu. Karmaşık bir hikâyeyi özetlemek istersek, Topluluk Standartları daha çok standartlara benzerken, Uygulama Standartları kurallara benziyordu.

Aynı şekilde, bir hava yolu şirketi pilotlarından ya kurallara ya da standartlara uymalarını isteyebilir. Soru pistte doksan dakika geçirdikten sonra kapıya geri dönüp dönmemek ya da emniyet kemeri işaretinin tam olarak ne zaman açılacağı olabilir. Hava yolu şirketi pilotların takdir yetkisini kısıtladığı ve dolayısıyla hatayı azalttığı için kuralları tercih edebilir. Ama aynı zamanda, bazı şartlarda, pilotların en iyi yargılarını kullanmasının daha

iyi olacağına da inanabilir. Bu durumlarda, standartlar, gürültü ürettiklerinde bile, kurallardan daha iyi olabilirler.

Bütün bu vakalarda ve başka örneklerde, kurallar ve standartlar arasında karar verenler gürültü sorununa, ön yargı sorununa ya da ikisine birden odaklanmalıdır. Büyük ya da küçük kuruluşlar bu kararı sürekli vermek zorundalar. Bazen bunu sezgisel olarak ve bir çerçeveye dayandırmadan yaparlar.

Standartlar birçok şekil ve boyutta gelir. Temelde tamamen içeriksiz olabilirler: "Şartlar altında, uygun olanı yap." Kurallara yaklaşacak şekilde yazılabilirler; örneğin, karar vericinin takdir yetkisini kısıtlamak için, neyin uygun olduğunun açık bir şekilde tanımlandığı durumlarda. Kurallar ve standartlar ayrıca karıştırılıp eşleştirilebilir. Örneğin, bir personel bürosu standardın uygulanmasından önce gelecek bir kural benimseyebilir. (Kural: "Başvuru sahipleri üniversite mezunu olmalı". Standart: "Bu kısıtlamaya tabi olarak, harika bir iş çıkaracak insanları seçin.")

Kuralların gürültüyü azaltması ya da mümkünse tamamen ortadan kaldırması gerektiğini, standartların ise genellikle bol miktarda gürültü üreteceğini söylemiştik (elbette bir gürültüyü azaltma stratejisi benimsenmemişse). Özel ve kamusal örgütlerde, gürültü genellikle bir kural koyma başarısızlığının ürünüdür. Gürültü yeterince yükselince –benzer durumdaki insanların benzer muamele görmediğini herkes görebilir hâle gelince– genellikle kurallara doğru bir kayma gerçekleşir. Cezai hükümlerde olduğu gibi, hareket bir haykırışa dönüşebilir. Tipik olarak o haykırışın öncesinde bir tür gürültü denetimi gerçekleşir.

Bastırılanın Geri Dönüşü

Önemli bir soruyu düşünün: Kimler, çalışamayacak durumda olan insanlara ayrılan ekonomik faydalardan yararlanmaya uygun bulunmak anlamında, engelli sayılır? Soru bu şekilde yöneltilirse,

yargıçlar bu amaca özel, gürültülü ve dolayısıyla haksız kararlar verirler. Amerika Birleşik Devletleri'nde bu tür gürültülü ve haksız kararlar bir zamanlar çok yaygındı ve sonuçları skandal düzeyinde olurdu. Tekerlekli sandalyeye mahkûm ya da ağır depresyon veya kronik ağrıdan muzdarip, görünürde benzer durumdaki insanlar farklı muameleler görürdü. Buna tepki olarak, kamu görevlileri kurala daha yakın bir uygulamaya yöneldiler: *engellilik matrisi*. Matris eğitim, coğrafi konum ve geri kalan fiziksel yeterliliklere dayanarak nispeten mekanik yargılar gerektiriyor. Amaç kararları daha az gürültülü kılmak.

Sorunun hukuk profesörü Jerry Mashaw tarafından kaleme alınan başlıca tartışması, gürültülü yargıları ortadan kaldırma çabasına bir ad koydu: *bürokratik adalet.* Terim hatırlanmaya eğer. Mashaw matrisin yaratılmasını, özelikle gürültüyü ortadan kaldırmayı vadettiği için, temelde adil bularak yüceltiyor. Ancak bazı durumlarda, bürokratik adalet vaadi gerçekleştirilemeyebiliyor. Bir kurum kurallara bağlı kararlara kaydığı anda, o gürültünün yeniden su yüzüne çıkma riski doğuyor.

Kuralların belli vakalarda korkunç sonuçlar ürettiğini varsayın. Böyle olursa, karar vericiler kuralları, aşırı sert olduklarını düşünerek açıkça yok sayabilirler. Bu nedenle sivil itaatsizliğin denetlenmesi ve hatta görülmesi zor olan hafif bir biçimi üzerinden, takdir yetkisi kullanabilirler. Özel şirketlerde, çalışanlar aptalca görünen katı kuralları yok sayarlar. Benzer şekilde, kamu güvenliğini ve sağlığını korumakla yükümlü idari birimler, çok katı ya da kural gibi olduklarında, mevzuatları uygulamayı açıkça reddedebilirler. Ceza hukukunda, jürinin *karar bozması*, jüri üyelerinin mantıksızca katı ve gaddar olduğu gerekçesiyle kanuna uymayı reddettiği durumlar anlamına gelir.

Kamusal ya da özel bir kurum gürültüyü katı kurallar aracılığıyla kontrol etmeye çalıştığı zaman kuralların takdir yetkisini yeraltına çekmesi olasılığına karşı her zaman uyanık olmalıdır.

Üç vuruş politikasında, savcıların sık sık ortaya koydukları tepkiyi –iki kez hüküm giymiş insanlara karşı suçlamada bulunmadan kaçınmak– kontrol etmek ve hatta görmek son derece zordu.

Böyle şeyler olduğu zaman, ortada gürültü olur ama kimse duymaz. Niyet edildiği gibi işlediklerinden emin olmak için kurallarımızı gözetim altında tutmalıyız. Eğer işlemiyorlarsa, gürültünün varlığı bir ipucu olabilir ve kurallar gözden geçirilmelidir.

Bir Çerçeve

İş dünyasında ve hükûmette, kurallar ve standartlar arasında tercih genellikle sezgisel olarak yapılır ama daha disiplinli olması sağlanabilir. İlk kabaca tahmin olarak, tercih sadece iki faktöre dayanır: (1) kararların maliyeti ve (2) hataların maliyeti.

Standartlarda, kararların maliyetleri, sadece içerik sağlamaları gerekeceği için, her tür yargıç için çok yüksek olabilir. Yargı üretmek zahmetli olabilir. Doktorlardan en iyi yargılarını üretmeleri istense, her vakayı düşünmeye vakit ayırmaları gerekebilir (ve yargılar pekâlâ gürültülü olabilir). Doktorlara hastalarında boğaz enfeksiyonu olup olmadığına karar vermeleri için açık kılavuzlar verilse, kararları hızlı ve nispeten direkt olabilir. Hız limiti saatte yüz kilometreyse, polis memurlarının insanların ne kadar hızlı gitmesine izin olduğunu düşünmesine gerek kalmaz ama eğer standart insanların "mantıksız şekilde hızlı" gidemeyeceğini söylerse, memurların daha fazla kafa yorması gerekebilir (ve uygulamanın gürültülü olacağı kesindir). Kurallarda, kararların maliyeti tipik olarak daha düşüktür.

Yine de, karmaşık bir durum söz konusu. Bir kez yürürlüğe konduklarında kuralların uygulanması dolambaçsız olabilir ama bir kural konmadan önce, *birinin ona karar vermesi gerekir.* Bir kural üretmek zor olabilir. Bazen yanına yaklaşılmayacak kadar maliyetli olur. Hukuk sistemleri ve özel şirketler, bu nedenle sık

sık *makul, ihtiyatlı ve yapılabilir* gibi sözcükler kullanırlar. Bu tür terimlerin tıp ve mühendislik alanlarında da eşit derecede önemli rol oynamasının bir nedeni de budur.

Hataların maliyeti, sayılarına ve büyüklüklerine bağlıdır. Her yerde karşımıza çıkan bir soru, temsilcilerin bilgili ve güvenilir olup olmadığı ve karar hijyeni uygulayıp uygulamadıklarıdır. Bilgili ve güvenilirlerse ve karar hijyeni uyguluyorlarsa, o zaman bir standart iyi iş çıkarabilir ve gürültü çok az olabilir. Temsilcilerine güvenmemek için nedenleri olduğunda, yöneticiler kurallar dayatmak ihtiyacı duyarlar. Temsilciler yetersiz ya da ön yargılıysa ve karar hijyenini uygulanabilir şekilde yerleştiremezlerse, o zaman kurallarla kısıtlanmaları gerekir. Sağduyulu örgütler tanıdıkları takdir yetkisinin, temsilcilerine duydukları güvenle yakından bağlantılı olduğunu bilirler.

Elbette kusursuz güvenden, tam güvensizliğe bir süreklilik vardır. Bir standart hiç güven uyandırmayan temsilcilerde sayısız hataya yol açabilir; ancak eğer hatalar küçük ve önemsizse, hoşgörülebilirler. Bir kural sadece çok az hataya yol açabilir ama eğer hatalar feci sonuçlara neden olursa, bir standardı tercih edebiliriz. Hataların maliyetinin kurallardan ya da standartlardan daha büyük olduğunu düşünmek için *genel* bir neden olmadığını görebiliyor olmalıyız. Elbette, bir kural kusursuzsa hiç hata üretmez. Ama kurallar nadiren kusursuzdurlar.

Hukukun ancak 21 yaşında ve daha büyükseniz alkol alabileceğinizi söylediğini varsayın. Hukuk genç insanları alkol tüketimiyle bağdaştırılan çeşitli risklerden korumayı hedefler. Bu şekilde anlaşıldığında, hukuk birçok hata üretecektir. 20, 19, 18 ve hatta 17 yaşındaki bazı insanlar alkolü rahatlıkla idare edebilirler. Ama 22, 42 ya da 62 yaşındaki bazı insanlar için aynısı söylenemez. En uygun sözcükleri bulabilseydik ve insanlar bu sözcükleri doğru şekilde uygulayabilseydi, bir standart daha az hata üretirdi. Elbette

bunu yapmak zordur ve alkol satışında neredeyse her zaman yaşa dayanan basit kurallar görmemiz bu yüzdendir.

Bu örnek çok daha büyük bir noktayı akla getiriyor. Çok sayıda karar alınması gerektiği zaman, çok gürültü olabilir ve net kuralların konması gerekliliği ağırlık kazanabilir. Dermatologlar kaşıntılı kızarıklığı ve benleri olan çok sayıda hasta görüyorlarsa, yargıları sağduyulu kurallarla sınırlandığında daha az hata yapabilirler. Bu tür kurallar olmadığında, açık uçlu standartlarla, kararların maliyeti akılalmaz derecede büyüyebilir. Tekrarlı kararlarda, duruma özel yargılar yerine mekanik kurallara yönelmenin gerçek avantajları vardır. Takdir yetkisi kullanmanın yükleri çok büyük olabilir ve gürültünün maliyetleri ya da yarattığı haksızlık katlanılmaz olabilir.

Akıllı örgütler her iki düzenleyici davranış yolunun da dezavantajlarının farkındadır. Gürültüyü (ve ön yargıyı) azaltmanın bir yolu olarak kurallar ya da kurala yakın standartlar koyarlar. Hataların maliyetlerini en aza indirmek adına, kuralların isabetli (yeterince) olduğundan emin olmak için, öncesinde zaman ve dikkat harcamaya da isteklidirler.

Gürültüyü Yasaklamak mı?

Pek çok durumda gürültü skandal olmalıdır. İnsanlar bununla yaşarlar ama bunu yapmak zorunda olmamalıdırlar. Gürültüye basit bir tepki açık uçlu takdir yetkisinden ya da muğlak bir standarttan bir kurala ya da ona yakın bir şeye kayma olacaktır. Artık, basit tepkinin ne zaman en doğru tepki olduğu konusunda bir hissimiz var. Ancak bir kural uygulanamaz olduğunda ya da iyi bir fikir olmadığında bile, gürültüyü azaltmak için çeşitli stratejiler sunduk.

Bütün bunlar akla büyük bir soruyu getiriyor: Yasal sistem gürültüyü yasaklamalı mı? Evet cevabını vermek çok kolay olurdu ama hukukun gürültüyü kontrol etmek için, şimdi yaptığından daha

fazlasını yapması gerek. Sorunu düşünmenin bir yolu şu olabilir. Alman sosyolog Max Weber genel kuralların disiplin altına sokmadığı gayriresmî ve duruma özel yargılar olarak anladığı "Kadı adaletinden" yakınır. Weber'e göre, Kadı adaleti hoş görülemez bir şekilde vakaya özgü oluyordu, hukuk kuralının açıkça çiğnenmesiydi. Weber'in ifadesiyle, Kadı "kesin olarak resmî kurallara göre ve 'kişilerden bağımsız' olarak hüküm vermiyordu. Büyük ölçüde tam tersi oluyordu, Kadı kişileri somut özelliklerine göre, somut durumun ışığında ve somut sonucun hakkaniyetine ve uygunluğuna göre yargılıyordu."

Weber bu yaklaşımın, "hiçbir akılcı karar kuralını tanımadığını" savunuyordu. Weber'in Kadı adaletinin neden olduğu dayanılmaz gürültüden yakındığını kolayca görebiliriz. Weber önceden disiplin altına sokulan bürokratik yargıların yükselişini yüceltiyordu (bürokratik adalet fikrini hatırlayın). Özelleşmiş, profesyonel ve kurala bağlı yaklaşımları hukukun evriminin nihai aşaması olarak görüyordu. Ancak Weber'in bunları yazmasından çok uzun zaman sonra bile, Kadı adaletinin ya da ona benzer bir şeyin yaygınlığını hâlâ koruduğunu görmek zor değil. Asıl soru, bu konuda ne yapılacağı.

Gürültüyü azaltmanın Evrensel İnsan Hakları Bildirisi'nin bir parçası olması gerektiğini söyleyecek kadar ileri gidemeyiz, ancak bazı durumlarda, gürültü bir hak çiğneme olarak görülebilir ve genel anlamda, dünyanın her yerindeki yasal sistemler gürültüyü kontrol etmek için çok daha büyük çabalar harcamalılar. Cezai hüküm konusunu, kabahatlere karşılık verilen para cezalarını, sığınma hakkının tanınması ya da tanınmamasını, eğitim fırsatlarını, vizeleri, inşaat izinlerini ve mesleki lisansları düşünün. Ya da büyük bir hükûmet bürosunun yüzlerce hatta binlerce insanı işe aldığını ve kararlarının bir ahenginin ya da nedeninin olmadığını, bir gürültü kakofonisi olduğunu düşünün. Ya da küçük çocuklara, dosyada hangi memurun görevlendirildiğine göre çok farklı dav-

ranan bir çocuk velayet şubesi düşünün. Bir çocuğun hayatının ve geleceğinin bu piyangoya bağlı olması ne kadar kabul edilebilir?

Pek çok örnekte, bu tür kararlardaki değişkenliği açıkça teşhis edilebilir bilişsel ön yargılar ve belli ayrımcılık biçimlerinin de aralarında olduğu ön yargılar yönlendirir. Böyle olduğunda, insanlar durumu katlanılmaz görürler ve yasa, yeni ve farklı uygulamaları talep eden bir düzeltici olarak yardıma çağrılabilir. Dünyanın her yerinde örgütler ön yargıyı kötü karakter olarak görüyor. Haklılar. Ama gürültüyü aynı şekilde görmüyorlar. Görmeliler.

Pek çok alanda, mevcut gürültü seviyesi çok yüksek. Gürültü çok yüksek maliyetler dayatıyor ve korkunç bir haksızlık üretiyor. Bizim burada aktardığımız kısım buz dağının görünen ucu. Hukuk bu maliyetleri azaltmak için çok daha fazlasını yapmalı. Bu haksızlıkla mücadele etmeli.

Kurallar ve Standartlardan Bahsederken

"Kurallar hayatı basitleştirir ve gürültüyü azaltır. Ama standartlar insanların kendilerini durumun özelliklerine göre ayarlamalarına izin verir."

"Kurallar mı, standartlar mı? Önce, hangisinin daha fazla hata ürettiğini sorun. Sonra hangisini üretmenin ve birlikte hareket etmenin daha kolay ve daha zahmetli olduğunu sorun."

"Genellikle açıkça gürültüye dikkat etmediğimiz için, kurallar benimsememiz gerekirken standartlar kullanırız."

"Gürültüyü azaltma Evrensel İnsan Hakları Bildirisi'nin bir parçası olmamalı, en azından şimdilik. Yine de gürültü son derece haksız olabilir. Dünyanın her yerindeki hukuk sistemleri gürültüyü azaltmak için güçlü adımlar atmayı düşünmeli."

GÖZDEN GEÇİRME VE SONUÇ

Gürültüyü Ciddiye Almak

Gürültü yargıların istenmeyen değişkenliğidir ve çok fazlasıyla karşı karşıyayız. Burada temel amaçlarımız neden böyle olduğu ve bu konuda ne yapılabileceğiydi. Bu kitapta materyalin büyük kısmını kapsadık ve sonuç olarak, ana noktaların kısa bir özetiyle, daha geniş bir bakış açısını sunmak istiyoruz.

Yargılar

Bizim kullandığımız hâliyle, *yargı* "düşünmek"le karıştırılmamalı. Bu, çok daha dar bir kapsam: Yargı aletin insan zihni olduğu bir ölçüm şeklidir. Diğer ölçümler gibi, yargı da bir nesneye puan verir. O puanın illa bir sayı olması gerekmez. "Mary Johnson'ın tümörü muhtemelen iyi huylu", tıpkı "Ulusal ekonomi çok istikrarsız", "Fred Williams yeni müdürümüz olarak işe alınacak en iyi insan olur", ve "Bu riskin sigorta primi 12.000 dolar olmalı", beyanları gibi bir yargıdır. Yargılar çeşitli bilgi parçalarını genel bir değerlendirmeye gayriresmî olarak entegre ederler. Hesaplama değildirler ve kesin kurallara uymazlar. Bir öğretmen bir kompo-

zisyona not verirken yargı kullanır ama çoktan seçmeli bir testi puanlarken yargı kullanmaz.

Pek çok insan hayatını profesyonel yargılar üreterek kazanır ve bu yargılardan herkes önemli şekillerde etkilenir. Burada bizim verdiğimiz isimle profesyonel *yargıçlar* arasında futbol koçları, kardiyologlar, avukatlar ve mühendisler, Hollywood yapımcıları, risk değerlendirme uzmanları ve daha pek çoğu sayılabilir. Profesyonel yargılar hem kapsamlı olarak incelendikleri hem de kalitelerinin hepimizin üstünde çok büyük etkisi olması nedeniyle) bu kitabın odağı oldular. Öğrendiklerimizin, insanların hayatlarının başka kısımlarında vardıkları yargılara da uyduğuna inanıyoruz.

Bazı yargılar *öngörücü*dür ve bazı öngörücü yargılar doğrulanabilir: İsabetli olup olmadıklarını önünde sonunda öğreniriz. Bu genellikle, bir ilacın etkileri, bir salgının gidişatı ya da bir seçimin sonuçları gibi, kısa vadeli sonuç tahminleri için geçerlidir. Ancak uzun vadeli tahminlerin ve kurgu soruların cevaplarının da aralarında olduğu pek çok yargı, doğrulanamaz. Bu tür yargıların kalitesi ancak onları üreten düşünce sürecinin kalitesiyle değerlendirilebilir. Dahası pek çok yargı öngörücü değil, *değerlendirmeci*dir: Bir yargıcın verdiği hüküm ya da bir tablonun ödüllü bir yarışmadaki sıralaması nesnel bir gerçek değerle kolay kolay karşılaştırılamaz.

Ancak çarpıcı bir şekilde, yargı üreten insanlar, olup olmamasına bakmaksızın gerçek bir değer varmış gibi davranırlar. Nişan alınacak görünmez –ve onların ve başkalarının çok fazla ıskalamaması gereken– bir boğa gözü varmış gibi düşünür ve hareket ederler. *Kişisel yargı* ifadesi hem anlaşmazlık olasılığını hem de sınırlı olacağı beklentisini ima ediyor. Yargı meseleleri bir *sınırlı anlaşmazlık* beklentisiyle tanımlanırlar. Görüş ayrılığına izin verilmeyen hesaplama meseleleriyle, aşırı uç örnekler dışında çok az anlaşma beklentisi olan zevk meseleleri arasında bir yer kaplarlar.

Hatalar: Ön Yargı ve Gürültü

Bir yargı grubunda hataların çoğu aynı istikamette olduğu zaman, *ön yargı* olduğunu söyleriz. Ön yargı, örneğin bir nişancı takımının atışlarının tutarlı olarak hedefin altına ve sol tarafına isabet ettiği, üst düzey yöneticilerin satışlar konusunda her sene aşırı iyimser olduğu ya da bir şirket üstünü çizmesi gereken başarısız projelere para yatırmaya devam ettiği zamanlarda oluşan *ortalama hata*dır.

Yargılardan ön yargıyı kaldırmak bütün hataları gidermez. Ön yargı giderildiğinde kalan hatalar ortak değildir. Yargıların istenmeyen uyumsuzluğu, gerçekliğe uyguladığımız ölçüm aletinin güvenilmezliğidir. *Gürültüdür.* Gürültü birbirinin eşi olması gereken yargılardaki değişkenliktir. Acil servisteki doktorlar, cezai hükümler veren yargıçlar ve sigorta şirketlerindeki risk değerlendirme uzmanları gibi, karar alıcı personelin kendi aralarında yer değiştirdiği örgütlerde gözlemlenen gürültü için *sistem gürültüsü* terimini kullanırız. Bu kitabın büyük kısmı sistem gürültüsüyle ilgiliydi.

Ön Yargıyı ve Gürültüyü Ölçmek

Ortalama karesel hata (OKH) iki yüzyıldır bilimsel ölçümlerdeki doğruluk standardıdır. OKH'nin ana özellikleri örnek ortalamayı nüfus ortalamasının ön yargısız bir tahmini olarak kullanması, olumlu ve olumsuz hataları eşit şekilde ele alması ve büyük hataları orantısız bir şekilde cezalandırmasıdır. OKH yargı hatalarının genellikle asimetrik olan gerçek maliyetlerini yansıtmaz. Ancak, profesyonel kararlar her zaman isabetli tahminleri gerektirir. Kasırgayla karşı karşıya olan bir şehir için, tehlikeyi aşırı küçük görmenin ya da büyütmenin maliyetleri elbette aynı değildir ama bu maliyetlerin meteoroloji uzmanlarının fırtınanın sürati ve yolu konusundaki tahminlerini etkilemesini istemezsiniz. OKH, hedefin

nesnel isabetlilik olduğu durumlarda, bu tür öngörücü yargılarda bulunmak için uygun standarttır.

OKH ile ölçüldüğü üzere, ön yargı ve gürültü bağımsız ve ek hata kaynaklarıdır. Tahmin edileceği gibi, ön yargı her zaman kötüdür ve azaltılması isabetliliği her zaman artırır. Gürültünün de eşit derecede kötü olması ve gürültüyü azaltmanın her zaman iyileştirme getirmesi ise daha az sezgiseldir. Yargılar açıkça ön yargılı olduğunda bile, en iyi saçılma miktarı sıfırdır. Amaç, elbette hem ön yargıyı hem gürültüyü en aza indirmektir.

Bir doğrulanabilir yargılar grubunda ön yargı bir vakanın ortalama yargısıyla, karşılık gelen gerçek değer arasındaki farktır. Bu karşılaştırma doğrulanamaz yargılar için imkânsızdır. Örneğin, bir risk değerlendirme uzmanının belli bir risk için belirlediği primin gerçek değeri hiçbir zaman bilinmeyecektir. Aynı şekilde, belli bir suç için adil cezanın gerçek değerini de kolayca bilemeyiz. Bu bilginin eksikliğinde, sık bulunulan uygun bir varsayım (her zaman doğru olmamakla birlikte) yargıların ön yargısız olduğu ve birden fazla yargıcın ortalamasının gerçek değerin en iyi tahmini olduğudur.

Bir sistemdeki gürültü, *gürültü denetimiyle* ölçülebilir. Gürültü denetimi çok sayıda profesyonelin aynı vakalar için (gerçek ya da kurgu) bağımsız yargılarda bulunduğu bir deneydir. Gürültüyü, tıpkı bir hedef tahtasının arkasına bakarak atışların saçılımını görebildiğimiz gibi, gerçek bir değer bilmeden ölçebiliriz. Gürültü denetimleri, bir radyoloji departmanı ya da adli yargı sistemi gibi pek çok sistemde yargıların çeşitliliğini ölçebilir. Bu denetimler zaman zaman beceri ya da eğitim eksiklerine dikkat çekebilirler. Ayrıca, örneğin aynı takımdaki risk değerlendirme uzmanları risk değerlendirmelerinde farklılık gösterdiklerinde, sistem gürültüsünün miktarını da saptayabilirler.

Ön yargı ve gürültüden hangisi daha büyük bir sorundur. Bu, duruma göre değişir. Cevap pekâlâ gürültü de olabilir. Ön yargı

ve gürültü, hataların (ön yargıların) ortalaması hatanın standart sapmalarına (gürültü) eşit olduğu zaman, toplam hataya (OKH) eşit katkılarda bulunurlar. Yargıların dağılımı normal olduğunda (standart çan eğrisi), yargıların %84'ü gerçek değerin üstünde (ya da altında) olduğu zaman, ön yargıların ve gürültünün etkisi eşittir. Bu, genellikle profesyonel bağlamda saptanması mümkün olabilen, önemli bir ön yargı demektir. Ön yargı bir standart sapmadan küçük olduğu zaman, toplam hatanın daha büyük kaynağı gürültüdür.

Gürültü Bir Sorundur

Değişkenlik bazı yargılarda sorun yaratmaz, hatta hoş karşılanır. Görüş farklılıkları fikir ve seçenek üretmede önemlidir. Aykırı düşünme yenilik için gereklidir. Film eleştirmenleri arasında görüşlerin çokluğu can sıkıcı değil, bir özelliktir. Tüccarlar arasındaki anlaşmazlıklar piyasayı oluşturur. Rekabet hâlindeki yeni şirketlerin strateji farklılıkları, piyasaların en uygun olanı seçmesine olanak sağlar. Ancak yargı meselesi dediğimiz şeylerde ise, sistem gürültüsü her zaman bir sorundur. İki doktor farklı teşhislerde bulunursa, en az biri hatalı demektir.

Bu kitabı motive eden sürprizler sistem gürültüsünün şaşırtıcı büyüklüğü ve verdiği hasarın çarpıcılığı oldu. Her ikisi de yaygın beklentilerin çok üstünde. İş dünyası, tıp, adli yargı sistemi, parmak izi analizi, tahminler, kişisel puanlamalar ve politika gibi pek çok alandan örnekler verdik. Çıkarımımız şudur: Yargının olduğu yerde gürültü vardır, hem de sandığınızdan çok daha fazla.

Gürültünün hatadaki büyük rolü, yaygın olarak sahip olduğumuz, rastgele hataların "birbirlerini iptal ettikleri için" önemli olmadığı inancıyla çelişir. Bu inanç yanlıştır. Çok sayıda atış hedef tahtasının her yerine yayılmışsa, ortalamada boğa gözüne isabet ettiklerini söylemek, işe yaramaz. Bir işe başvuran aday hak

ettiğinden daha yüksek, diğeri hak ettiğinden daha düşük puan alırsa, yanlış kişi işe alınabilir. Bir sigorta poliçesi aşırı fiyatlanır, diğerinin fiyatı çok düşük tutulursa, iki hatanın da şirkete maliyeti olur: Biri iş, diğeri para kaybettirir.

Kısaca, yargılar geçerli bir neden olmadan değişkenlik gösterdiğinde, hata olduğundan emin olabiliriz. Gürültü yargıların doğrulanamadığı ve hatanın ölçülemediği durumlarda bile yıkıcıdır. Benzer durumlardaki insanların farklı muamele görmesi haksızlıktır ve profesyonel yargıların tutarsız olarak görüldüğü bir sistem inandırıcılığını kaybeder.

Gürültü Çeşitleri

Sistem gürültüsü *yatay gürültü* ve *örüntü gürültüsü* olarak parçalara ayrılabilir. Bazı yargıçlar diğerlerinden genel olarak daha katıdır ve diğerleri daha merhametlidir. Bazı tahminciler piyasa beklentileri konusunda genellikle fazla iyimserdir, diğerleri kötümser. Kimi doktorlar diğerlerinden daha fazla antibiyotik reçete eder. *Yatay gürültü* farklı bireylerin ortalama yargılarının değişkenliğidir. Yargı ölçeklerinin muğlaklığı yatay gürültü kaynaklarından biridir. *Olası* gibi sözcükler ya da sayılar (örneğin, "0 ile 6 arası bir ölçekte 4" gibi) farklı insanlara farklı şeyler ifade eder. Yatay gürültü yargı sistemlerinde önemli bir hata kaynağı ve gürültüyü azaltmayı amaçlayan müdahaleler için önemli bir hedeftir.

Sistem gürültüsü (ve genellikle daha büyük) bir ögeyi daha içerir. Yargılarının ortalamasından bağımsız olarak, iki yargıç hangi suçların daha sert cezaları hak ettiği konusunda farklı görüşlere sahip olabilirler. Hüküm kararları vakaların sıralamadaki yerlerinde farklılığa sebep olur. Bu değişkenliğe *örüntü gürültüsü* (teknik terim: *istatistiksel etkileşim)* diyoruz.

Örüntü gürültüsünün ana kaynağı sabittir: Yargıçların aynı vakaya kişisel ve kendilerine has tepkilerinin farkıdır. Bu farkla-

rın bir kısmı bireylerin bilinçli ya da bilinçsiz olarak uydukları prensipler ve değerleri yansıtır. Örneğin, bir yargıç dükkân hırsızlarına karşı özellikle sertken, trafik suçlularına karşı alışılmadık düzeyde merhametli olabilir ve bir diğer yargıç tam tersi bir kalıp sergileyebilir. Altta yatan prensiplerin ve değerlerin bazıları hayli karmaşık olabilir ve yargıç onların farkında olmayabilir. Mesela bir yargıç daha ileri yaştaki dükkân hırsızlarına farkında olmadan daha yumuşak davranabilir. Son olarak belli bir vakaya hayli kişisel bir tepki de sabit olabilir. Bir yargıcın kızına benzeyen bir sanık, başka bir gün, aynı sempati hissini, dolayısıyla merhameti uyandırabilir.

Sabit örüntü gürültüsü yargıçların kendilerine özgülüğünü yansıtır: Vakalara tepkileri kişilikleri kadar bireyseldir. İnsanlar arasındaki ince farklılıklar genellikle keyifli ve ilginçtir ama profesyoneller tutarlılık gerektiren bir sistemin parçası olduklarında farklılıklar soruna dönüşür. İncelediğimiz çalışmalarda, bu tür bireysel farkların ürettiği sabit örüntü gürültüsü genellikle sistem gürültüsünün en büyük kaynağıdır.

Bununla birlikte, yargıçların belli vakalara karşı farklı tutumları da tamamen sabit değildir. Örüntü gürültüsünün, *durum gürültüsü* denen geçici bir ögesi vardır. Bu tür bir gürültüyü, bir radyoloğun aynı görüntüye farklı günlerde farklı teşhisler yakıştırması ya da bir parmak izi uzmanının iki izi bir durumda eşleştirirken, diğerinde eşleştirmemesi hâlinde saptarız. Bütün bu örneklerin gösterdiği gibi, durum gürültüsü en kolay, bir yargıç önündeki vakayı daha önce gördüğü bir vaka olarak tanımadığı zaman ölçülür. Durum gürültüsünü ortaya koymanın bir başka yolu, bağlamın alakasız bir özelliğinin yargılar üstündeki etkisini göstermektir. Yargıçların tuttukları futbol takımı galibiyet aldığı zaman daha yumuşak olması ya da doktorların öğleden sonra daha fazla opioid içeren ilaç reçete etmesi gibi.

Yargının Psikolojisi ve Gürültü

Öngörücü yargılarda hata sebebi, yargıçların bilişsel kusurları değildir. *Objektif bilgisizlik* genellikle daha büyük bir rol oynar. Bazı olgular gerçekten bilinemez: Dün doğan bir bebeğin bundan yetmiş yıl sonra kaç torununun olacağı ya da önümüzdeki yıl yapılacak bir kurada kaç bilete piyango isabet edeceği gibi. Diğerleri belki bilinebilir ama yargıç bilmiyordur. İnsanların öngörücü yargıları konusunda abartılı öz güvenleri ön yargıları kadar objektif bilgisizliklerini de hafife alır.

Tahminlerimizin isabetliliğinin bir sınırı vardır ve bu sınır genellikle hayli düşüktür. Bununla birlikte yargılarımız konusunda genellikle rahatızdır. Bu tatmin edici öz güveni veren, olguları ve yargıyı tutarlı bir hikâyeye sığdırmanın sonucu kendimiz ürettiğimiz bir ödül olan *iç sinyaldir.* Yargımıza öznel güvenimiz illa yargının nesnel isabetliliğiyle alakalı değildir.

Öngörücü yargılarının isabetliliğinin sadece düşük değil, aynı zamanda formüllerinkinin altında olduğunu duymak pek çok insanı şaşırtır. Sınırlı veri üstüne kurulu basit doğrusal modeller ya da bir zarfın arka yüzüne karalanabilecek kurallar bile, insan yargıçları tutarlı olarak alt eder. Kuralların ve modellerin kritik avantajı gürültüsüz olmalarıdır. Bizim öznel olarak tecrübe ettiğimiz gibi, yargı incelikli ve karmaşık bir süreçtir, o inceliğin büyük kısmının gürültü olabileceğine dair bir göstergemiz yoktur. Basit kurallara körü körüne uymanın sıklıkla bizden daha fazla isabet üretebileceğini hayal etmek bizim için zordur ama bu artık defalarca kanıtlanmış bir olgudur.

Psikolojik ön yargılar, elbette bir sistematik hata kaynağıdır ya da istatistiksel ön yargıdır. Aynı zamanda gürültü kaynağı da oldukları o kadar belirgin değildir. Ön yargılar bütün yargıçlarda ortak olmadığı, farklı derecelerde var olduğu ve etkileri dış şartlara bağlı olduğu zaman, psikolojik ön yargılar gürültü üretirler. Örneğin, eğer işe alım kararlarını veren yöneticilerin yarısı kadınlara

karşı, yarısı kadınlar lehine ön yargılıysa, toplam bir ön yargı olmayacaktır ama sistem gürültüsü birçok işe alım hatasına yol açacaktır. Bir diğer örnek de ilk izlenimlerin orantısız etkisidir. Bu bir psikolojik ön yargıdır ama bu ön yargı, kanıtların sunum sırası rastgele değiştiğinde, durum gürültüsü üretir.

Yargı sürecini bir ölçek üzerinde bir yargı üretmek için bir işaret grubunun gayriresmî entegrasyonu olarak tanımladık. Sistem gürültüsü, dolayısıyla, yargıçların işaretleri kullanımlarında, işaretlere verdikleri ağırlıkta ve ölçeği kullanmalarında benzerliği korumalarını gerektirir. Durum gürültüsünün rastgele etkilerini bir kenara bırakınca bile, bu şartlar nadiren yerine getirilir.

Tek boyutlu yargılarda görüş birliği genellikle hayli yüksektir. Farklı işe alım uzmanları, iki adaydan hangisinin daha karizmatik ya da çalışkan olduğu konusundaki değerlendirmelerinde, genellikle aynı görüşü paylaşırlar. Ortak bir sezgisel süreç olan, yoğunluk boyutları arasında *eşleme* süreci –insanların yüksek not ortalamasını okuma yaşının çok erken olmasıyla eşleştirmesi gibi– genellikle benzer yargılar üretir. Aynı şey aynı genel istikameti işaret eden az sayıdaki işarete dayalı yargılar için de geçerlidir.

Büyük bireysel farklar, bir yargı *çoklu, çelişkili işaretleri tartmayı* gerektirdiği zaman su yüzüne çıkar. Aynı adaya bakarken, bazı işe alım uzmanları zekâ ya da karizma kanıtlarına daha fazla ağırlık verirken, diğerleri çalışkanlık ve baskı altında sakinlik gibi endişelerden daha fazla etkileneceklerdir. İşaretler tutarsız olduğunda ve tutarlı bir hikâyeye uymadıklarında, farklı insanlar kaçınılmaz olarak bazı işaretlere daha fazla ağırlık verip diğerlerini yok sayar. Ortaya örüntü gürültüsü çıkar.

Gürültünün Karanlığı

Gürültü öne çıkan bir sorun değildir. Nadiren tartışılır ve ön yargıdan daha az dikkat çekicidir. Muhtemelen siz de çok fazla kafa

yormamışsınızdır. Önemi düşünüldüğünde, gürültünün gölgede kalması başlı başına ilginç bir fenomendir.

Genellikle kötü yargıların açıklaması olarak bilişsel ön yargılar ve düşünmenin diğer duygusal ve güdülenmiş çarpıtmaları kullanılır. Analizciler sonu kötüye çıkan kararları izah etmek, aşırı öz güven, çapalama, kayıptan kaçınma, ulaşılabilirlik ön yargısı ve diğer ön yargılardan yardım alırlar. Ön yargıya dayalı bu açıklamalar tatmin edicidir çünkü insan zihni nedensel açıklamalar için yanıp tutuşur. Bir şey yolunda gitmediği anda, bir neden arar ve genellikle de buluruz. Pek çok durumda neden bir ön yargı gibi görünecektir.

Ön yargının kendine ait, gürültüde olmayan, bir açıklayıcı karizması var. Geriye dönüp bakarak bir kararın neden yanlış olduğunu açıklamaya çalışırsak, kolayca ön yargı bulur ama asla gürültü bulmayız. Gürültüyü görmemize sadece dünyaya *istatistiksel bir bakış* izin verir ama bu bakış kendiliğinden gelmez; neden hikâyelerini yeğleriz. Kurumlarımızda istatistiksel düşüncenin eksikliği gürültünün ön yargıdan çok daha az dikkat çekmesinin nedenlerinden biridir.

Bir diğer neden, profesyonellerin kendi yargılarında ve meslektaşlarınınkilerde gürültüyle yüzleşme ihtiyacını nadiren duymalarıdır. Bir eğitim döneminin ardından, profesyoneller genellikle tek başlarına kararlar alırlar. Parmak izi uzmanları, tecrübeli risk değerlendirme uzmanları ve kıdemli patent memurları, meslektaşlarının onlara hangi açılardan katılmayacağını düşünmeye nadiren zaman ayırırlar. Kendi kendileriyle nasıl ters düşebileceklerini düşünmeye ise çok daha az.

Çoğu zaman, profesyoneller kendi kararlarından emindirler. Meslektaşlarının onlarla hemfikir olmasını bekler ve olup olmadıklarını hiçbir zaman öğrenmezler. Çoğu alanda, bir yargı hiçbir zaman gerçek bir değerle karşılaştırılmayabilir ve en fazla, *saygı uyandıran uzman* kabul edilen bir diğer profesyonelin onayına

sunulur. Profesyoneller sadece zaman zaman şaşırtıcı bir anlaşmazlıkla karşılaşırlar ve bu olduğu zaman, genellikle durumu izole bir vaka olarak görmek için nedenler bulurlar. Örgütlerin rutinleri de bünyelerindeki uzmanlar arasında uyuşmazlık kanıtlarını yok sayma ya da bastırma eğilimindedir. Bu, anlaşılabilir. Örgütsel bir bakış açısıyla, gürültü bir utanç sebebidir.

Gürültü Nasıl Azaltılır (ve Tabii Ön Yargı da)?

Bazı insanların diğerlerinden daha iyi yargılar ürettiğine inanmak için nedenler çok. En iyi yargıçların başlıca özellikleri, göreve özel beceri, zekâ ve belli bir bilişsel tarzdır (en iyi aktif olarak açık görüşlü olmak olarak tanımlanır). Şaşırtıcı olmayan bir şekilde, iyi yargıçlar korkunç hatalar yapmazlar. Ancak bireysel farkların çoklu kaynakları düşünüldüğünde, en iyi yargıçların bile karmaşık yargı sorunlarında kusursuz bir görüş birliğine sahip olmalarını beklememeliyiz. Her birimizi benzersiz kılan eğitimlerin, geçmişin, kişiliklerin ve tecrübelerin sonsuz çeşitliliği, gürültüyü de kaçınılmaz kılar.

Hatayı azaltmak için bir strateji ön yargıyı gidermektir. Tipik olarak, insanlar yargılarındaki ön yargıyı, ya yargılarını olaydan sonra düzelterek ya da ön yargılarını, yargılarını etkilemeden önce terbiye ederek gidermeye çalışırlar. Biz bir grup ortamında verilen kararlarla özellikle uygulanabilir olan üçüncü bir seçenek sunuyoruz: ön yargıları gerçek zamanda, ön yargı işaretlerini teşhis edecek bir *karar gözlemcisi* görevlendirerek saptamak (Bakınız Ek B).

Yargıda gürültüyü azaltmak için başlıca önerimiz, *karar hijyeni.* Bu terimi seçtik, çünkü gürültünün azaltılması, sağlıktaki hijyen gibi, tanımlanamayan bir düşmana karşı önlem almaktır. Örneğin el yıkamak, bilinmeyen patojenlerin vücuduna girmesini önler. Aynı şekilde karar hijyeni de hataları ne olduklarını bilmeden önleyecektir. Karar hijyeni en az ismi kadar sevimsiz ve

öngörülebilir ön yargılarla zaferle sonuçlanacak bir mücadeleden daha az heyecan vericidir. Tanımlanamayan bir zararı önlenmenin gösterişi olmayabilir ama kesinlikle yapmaya değer.

Bir örgütte gürültüyü azaltma çabası her zaman bir gürültü denetimiyle başlamalıdır (bakınız Ek A). Denetimin önemli bir işlevi organizasyonun gürültüyü ciddiye alma taahhüdüne girmesini sağlamaktır. Temel bir getirisi, ayrı gürültü türlerinin değerlendirilmesidir.

Çeşitli alanlarda gürültüyü azaltma çabalarının başarılarından ve sınırlarından bahsettik. Şimdi karar hijyenini tanımlayan altı prensibi bir kez daha özetleyecek ve gürültüye neden olan psikolojik mekanizmalara nasıl hitap ettiklerini ve değindiğimiz belli karar hijyeni teknikleriyle ilişkilerini anlatacağız.

Yargının amacı bireysel ifade değil, isabetliliktir. Bu cümle yargıda karar hijyeninin ilk prensibi için adayımız. Bu kitapta yargı için kullandığımız dar ve spesifik tanımı yansıtıyor. Sabit örüntü gürültüsünün sistem gürültüsünün büyük bir ögesi olduğunu ve insanları aynı sorunda farklı görüşler oluşturmaya iten bireysel farkların ve yargı kişiliklerinin direkt sonucu olduğunu gösterdik. Bu gözlem bizi kaçılmaz olduğu kadar itici de gelecek bir sonuca çıkarıyor: Yargı bireyselliğinizi ifade edeceğiniz yer değildir.

Kişisel değerler, bireysellik ve yaratıcılık, düşünmenin ve karar almanın, hedef seçiminin, bir soruna yaklaşmanın yeni yollarını oluşturmanın ve seçenekler üretmenin de aralarında olduğu birçok safhası için gerekli, hatta esastır. Ancak sıra bu seçenekler hakkında bir yargıya varmaya gelince, bireysellik ifadeleri bir gürültü kaynağı olur. Amaç isabetlilik olduğunda ve başkalarının sizinle hemfikir olmasını beklediğinizde, diğer hünerli yargıçların sizin yerinizde olsalar ne düşüneceklerini de göz önünde bulundurmalısınız.

Bu prensibin radikal bir uygulaması yargının yerine kural ve algoritmaların konmasıdır. Algoritmaya dayalı değerlendirmenin gürültüyü ortadan kaldırması garantidir; hatta gürültüyü tama-

men ortadan kaldırabilecek tek yaklaşımdır. Algoritmalar pek çok önemli alanda zaten kullanımda ve rolleri gittikçe artıyor. Ama algoritmaların önemli kararların nihai aşamasında insan yargısının yerini alması olası değildir ve bunu iyi haber olarak görüyoruz. Ancak, yargı hem algoritmaların gerektiği şekilde kullanılmasıyla hem de kararları bir profesyonelin şahsına has özelliklere daha az bağımlı kılan yaklaşımların benimsenmesiyle iyileştirilebilir. Örneğin, karar kılavuzlarının yargıçların takdir yetkisini kısıtlayarak ya da doktorların teşhislerinde homojenliği destekleyerek gürültüyü azaltmaya ve kararları iyileştirmeye nasıl yardımcı olduğunu gördük.

İstatistiksel düşün ve vakaya dış bakış takın. Yargıçların bir davaya onu tekil bir sorundan çok, benzer davaların referans sınıfının bir üyesi olarak kabul ettiğinde, dış bakış takındığını belirtmiştik. Bu yaklaşım sadece eldeki vakaya dayanan ve onu nedensel bir hikâyenin içine gömen varsayılan düşünme modundan farklılık gösterir. İnsanlar vaka için benzersiz bir görüş biçimlendirmek için kendilerine has tecrübeleri kullandığı zaman, sonuç örüntü gürültüsü olur. Dış bakış bu soruna bir çaredir: Aynı referans sınıfını paylaşan profesyoneller daha az gürültülü olur. Ek olarak, dış bakış sık sık değerli öngörüler üretir.

Dış bakış prensibi tahminlerin benzer vakaların istatistiklerine dayanmasını destekler. Ayrıca tahminlerin orta karar olması tavsiyesine yol açar (teknik terim *regresiftir*; bakınız Ek C*).* Geniş geçmiş sonuçlar yelpazesine ve sınırlı tahmin edilebilirliklerine dikkat etmek, karar alıcıların kendi yargılarına duydukları güveni kalibre etmelerine yardımcı olmalıdır. İnsanlar öngörülemez olanı tahmin edemedikleri için kabahatli olamazlar ama öngörü tevazusundan yoksun oldukları için suçlanabilirler.

Yargıları çok sayıda bağımsız göreve ayırarak yapılandırın. Bu parçala ve fethet prensibini gerekli kılan, insanların var olan ya da yeni ortaya çıkan bir hikâyeye uymayan bilgiyi çarpıtmasına

ya da yok saymasına neden olan ve bizim *aşırı tutarlılık* olarak tanımladığımız psikolojik mekanizmadır. Bir vakanın farklı yönlerinin izlenimleri birbirini kirlettiği zaman, toplam isabetlilik zarar görür. Benzetme olarak, birbirleriyle iletişim kurmalarına izin verilen bir grup tanığın kanıt değerine ne olduğunu düşünün.

İnsanlar yargı sorununu bir dizi daha küçük göreve bölerek aşırı tutarlılığı azaltabilirler. Bu teknik, mülakatçıların tek seferde bir özelliği değerlendirdiği ve bir sonrakine geçmeden önce o özelliği puanladığı yapılandırılmış mülakatlar uygulamasına benzer. Yapılandırma prensibi Apgar puanı gibi teşhis kılavuzlarına ilham olur. Ayrıca, *aracı değerlendirmeler protokolü* olarak adlandırdığımız yaklaşımın tam kalbinde yer alır. Bu protokol karmaşık bir yargıyı çoklu olguya dayalı değerlendirmelere parçalar ve her birinin bağımsız olarak değerlendirildiğini garantilemeyi hedefler. Mümkün olan her durumda, bağımsızlık farklı takımlara değerlendirme görevi vererek ve aralarındaki iletişimi en alt seviyeye çekerek korunur.

Henüz olgunlaşmamış, erken sezgilere karşı koyun. Karar alıcılara yargıları konusunda güven veren yargıyı tamamlama iç sinyalini tarif etmiştik. Karar alıcıların bu ödüllendirici sinyalden vazgeçmeyi istememesi kılavuzların, algoritmaların ve ellerini bağlayan diğer kuralların kullanılmasına direnmelerinin başlıca nedenidir. Karar alıcılar yaptıkları tercih konusunda rahat olmaya ve sezgisel öz güvenin ödüllendirici hissine ulaşmaya açıkça ihtiyaç duyarlar. Ancak bu ödülü kendilerine yeterince olgunlaşmadan, erkenden vermemeliler. Kanıtın dengeli ve dikkatli değerlendirmesinden beslenen sezgisel bir tercih hızlı bir şekilde varılan bir yargıdan çok daha üstündür. Sezgi yasaklanmamalı ama bilgiyle beslenmeli, disipline edilmeli ve geciktirilmelidir.

Bu prensip *bilgiyi sıralama* tavsiyemize ilham veriyor: Yargı üreten profesyonellere söz konusu bilgi doğru bile olsa, ihtiyaç duymadıkları ve ön yargıya neden olabilecek bilgiler verilmemeli.

Örneğin adli tıpta, uzmanları bir şüpheyle ilgili diğer bilgilerden habersiz tutmak iyi bir uygulamadır. Aracı değerlendirmelerin kilit bir ögesi olan tartışma gündemlerinin kontrolü de buraya aittir. Verimli bir gündem sorunun farklı yönlerinin ayrı ayrı değerlendirilmesini ve bütüncül bir yargının değerlendirmelerin profili tamamlanana kadar geciktirilmesini sağlayacaktır.

Birden fazla yargıçtan bağımsız yargıları edinin ve sonra o yargıları bir araya toplamayı düşünün. Bağımsızlık şartı örgütlerin prosedürlerinde, özellikle katılımcıların görüşlerinin başkaları tarafından şekillendirildiği toplantılarda rutin olarak çiğnenmektedir. Şelale etkileri ve grup kutuplaşması nedeniyle, grup tartışmaları genellikle gürültüyü artırır. Katılımcıların yargılarını tartışmadan *önce* toplama prosedürü hem gürültünün boyutunu açığa çıkarır hem de yapıcı bir farklılıklar çözümlemesini kolaylaştırır.

Bağımsız yargıların ortalamasını almanın sistem gürültüsünü azaltması garantidir (ama ön yargıyı azaltmaz). Tekil bir yargı bütün olası yargılar topluluğunun arasından seçilmiş tek bir örnektir ve örnek boyutunu büyütmek tahminlerin kesinliğini artırır. Ortalama almanın avantajı yargıçların farklı becerileri ve tamamlayıcı yargı örüntüleri olduğu zaman daha da güçlenir. Gürültülü bir grubun ortalaması oy birliğiyle varılan bir yargıdan daha isabetli olabilir.

Tercihinizi göreceli yargılardan ve göreceli ölçeklerden yana kullanın. Göreceli yargılar mutlak yargılardan daha gürültülüdür çünkü nesneleri bir ölçekte sınıflandırma becerimiz sınırlıyken, ikili karşılaştırmalar yapma becerimiz çok daha iyidir. Karşılaştırmaları gerektiren yargı ölçekleri mutlak yargı gerektiren ölçeklerden daha az gürültülü olacaktır. Örneğin, bir *vaka ölçeği*, yargıçların vakaları herkesin aşina olduğu örneklerle tanımlanan bir ölçeğe yerleştirmesini gerektirir.

Az önce listelediğimiz karar hijyeni prensipleri sadece tekrarlı yargılara değil, tek seferlik ya da bizim *tekil* olarak adlandırdığımız önemli kararlara da uygulanabilir. Tekil kararlarda gürültünün varlığı, mantığa ters gelebilir: Tanım gereği, sadece bir kez karar vereceksiniz ölçülecek bir değişkenlik olmaz. Yine de gürültü vardır ve hatalara neden olur. Bir atıcı takımındaki gürültü sadece ilk atıcıyı eylem hâlinde görürsek, görünmez ama diğer atıcıları da görürsek saçılma belirginleşir. Benzer bir şekilde, tekil yargıları düşünmenin en iyi yolu onları sadece *bir kez varılan tekrarlı yargılar* olarak ele almaktır. İşte bu yüzden, karar hijyeni onları da iyileştirmelidir.

Karar hijyeni uygulamak nankör olabilir. Gürültü görünmeyen bir düşmandır ve görünmeyen bir düşmana karşı kazanılan zafer sadece görünmez bir zafer olabilir. Ama fiziksel sağlık hijyeninde olduğu gibi, karar hijyeni de hayati önem taşır. Başarılı bir operasyondan sonra, hayatınızı kurtaranın cerrahın becerisi olduğuna inanmak istersiniz –ve elbette öyledir– ama eğer cerrah ve ameliyathanedeki bütün personel ellerini yıkamazlarsa ölebilirsiniz. Hijyenden kazanılacak çok fazla itibar olmayabilir ama sonuçları olduğu kesindir.

Ne Kadar Gürültü?

Elbette gürültüyle mücadele karar alıcıların ve örgütlerin üstünde durdukları tek şey değildir. Gürültünün azaltılması çok maliyetli olabilir: Bir lise notlamadaki gürültüyü her sınav kâğıdını beş ayrı öğretmene okutarak ortadan kaldırabilir ama artan yük pek makul değildir. Uygulamada bir miktar gürültü kaçınılmaz olabilir; her vakayı ayrı ayrı değerlendiren, insanlara bir makinenin dişlileri gibi davranmayan ve karar alıcılara bir tür yetki hissi tanıyan bir yargı sisteminin gerekli bir yan etkisidir. Hatta bir miktar gürültü arzu edilebilir bile; eğer yarattığı çeşitlilik söz konusu sistemin zaman

içinde uyarlanmasını sağlarsa. Mesela gürültü değişen değer ve amaçları yansıtır ve pratikte ya da kanunda değişikliğe yol açan bir tartışmayı tetiklerse.

Belki de en önemlisi, gürültüyü azaltma stratejilerinin kabul edilemez eksileri olabilir. Algoritmalarla ilgili pek çok kaygı abartılıdır ama bazıları meşrudur. Algoritmalar bir insanın asla yapmayacağı aptalca hatalar üretebilir ve dolayısıyla, insanların yaptığı birçok hatayı önlemeyi başarsalar bile, inandırıcılıklarını kaybedebilirler. Yetersiz tasarım nedeniyle ya da yetersiz veriyle beslenmeleri sonucu ön yargılı olabilirler. Bir çehrelerinin olmaması güvensizlik uyandırabilir. Karar hijyeni uygulamalarının da kendi olumsuzlukları vardır: Kötü yönetilirlerse, kararları bürokratikleştirme ve otonomilerinin altının oyulduğunu hisseden personelin moralini bozma riskini taşırlar.

Bütün bu riskler ve sınırlamalar tam anlamıyla değerlendirilmeyi hak eder. Ancak, gürültüyü azaltmaya itirazın mantıklı olup olmadığı tartışma konusu olan gürültüyü azaltma stratejisine bağlıdır. Yargıları bir araya toplamaya –belki de çok maliyetli olduğu gerekçesiyle– yapılan bir itiraz, kılavuzların kullanımına uymayabilir. Daha net olması için, gürültüyü azaltmanın maliyeti getirilerini aştığı zaman, peşine düşülmemelidir. Maliyet-getiri hesaplaması, sıfır olmayan optimal bir gürültü seviyesini açığa çıkarabilir. Sorun şu ki gürültü denetiminin yokluğunda, insanlar yargılarında ne kadar gürültü olduğundan habersizdirler. Durum bu olunca, gürültüyü azaltmanın zorluğundan dem vurmak, gürültüyü ölçmemek için bir bahaneden başka bir şey değildir.

Ön yargı hatalara ve haksızlıklara yol açar. Gürültü de öyle ve buna rağmen, gürültü hakkında çok daha az şey yaparız. Yargı hatası rastgele olduğunda onu bir nedene bağlayabildiğimiz duruma göre daha hoş görülebilir ama daha az zarar verici değildir. Önemli olan şeyler hakkında daha iyi kararlar istiyorsak, gürültünün azaltılmasını ciddiye almalıyız.

SON SÖZ

Daha Az Gürültülü Bir Dünya

Düşünün, örgütler gürültüyü azaltacak şekilde yeniden tasarlansalardı, nasıl olurlardı? Hastaneler, işe alım komiteleri, ekonomi tahmincileri, hükûmet büroları, sigorta şirketleri, kamu sağlığı yetkilileri, adli yargı sistemleri, hukuk şirketleri ve üniversiteler, gürültü sorununa karşı uyanık olur ve azaltmak için çabalarlardı. Gürültü denetimleri rutin olurdu ve her yıl yapılabilirlerdi.

Örgütlerin liderleri ya insan yargısının yerine koymak ya da bugün olduğundan çok daha fazla alanda destek vermek için algoritmaları kullanırlardı. İnsanlar karmaşık yargıları daha basit aracı değerlendirmelere parçalardı. Karar hijyenini bilir ve önerilerine uyarlardı. Bağımsız yargılar ortaya çıkarılır ve bir araya toplanırdı. Toplantılar çok farklı görünür, tartışmalar yapılandırılırdı. Dış bakış karar sürecine daha sistemli bir şekilde dâhil edilirdi. Aleni anlaşmazlıklara daha sık rastlanır ve daha yapıcı bir şekilde çözülürlerdi.

Sonuç daha az gürültülü bir dünya olurdu. Çok daha fazla para tasarrufu yapılır, kamu güvenliği ve sağlığı iyileşir, hakkaniyet artar ve kaçınılabilir birçok hata önlenirdi. Bu kitabı yazmaktaki amacımız bu fırsata dikkat çekmekti. Sizlerin bu fırsatı değerlendireceğinizi umuyoruz.

EK A

Gürültü Denetimi Nasıl Yapılır?

Bu ek gürültü denetimi için pratik bir kılavuz sağlıyor. Bir örgüt tarafından çalışanlarının ürettiği profesyonel yargıların kalitesini birimlerden birinde bir gürültü denetimi gerçekleştirerek incelemek için tutulmuş bir danışmanın bakış açısıyla okumalısınız.

Adından da anlaşılacağı gibi, bu denetimin odağı gürültünün yaygınlığıdır. Ancak iyi yürütülen bir denetim ön yargılar, kör noktalar, çalışanların eğitimindeki belirli eksikler ve çalışmalarının denetlenmesi konusunda da değerli bilgiler sağlayacaktır. Başarılı bir denetim, birimin operasyonlarında, profesyonellerin yargılarını yönlendiren doktrin, aldıkları eğitim, yargılarını desteklemek için kullandıkları araçlar ve çalışmalarının rutin denetimi de dâhil olmak üzere değişikliklere ilham vermelidir.

Bir gürültü denetimi hatırı sayılır miktarda çalışma ve detaya dikkati gerektirir çünkü bulguları anlamlı kusurları açığa çıkarırsa, inandırıcılığı elbette sorgulanacaktır. Vakaların ve prosedürün her detayı, bu nedenle, düşmanca ve detaylı bir inceleme beklentisiyle değerlendirilmelidir. Tarif edeceğimiz süreç, muhalefeti denetimin

potansiyel eleştirmeni olabilecek profesyonelleri denetçi olarak atayarak azaltmayı amaçlamaktadır.

Danışmanın (içeriden ya da dışarıdan olabilir) yanı sıra, alakalı karakterler aşağıdaki kişileri içerir.

Proje ekibi. Proje ekibi çalışmanın bütün aşamalarından sorumludur. Danışmanlar içeridense, proje ekibinin çekirdeğini onlar oluşturur. Ancak eğer danışmanlar dışarıdansa, bir iç proje ekibi onlarla omuz omuza çalışır. Bu, şirketteki insanların denetimi *kendi* projeleri, danışmanları da destek rol üstlenmiş kişiler olarak görmelerini sağlar. Verilerin toplanmasına, sonuçların analiz edilmesine ve nihai bir raporun hazırlanmasına liderlik eden danışmanlara ek olarak, proje ekibi yargıçların değerlendireceği vakaları kuracak konu uzmanlarını içermelidir. Proje ekibinin bütün üyelerinin yüksek profesyonel güvenilirliği olmalıdır.

Müşteriler. Bir gürültü denetimi ancak önemli değişimlere yol açarsa işe yarar ve bu, projenin "müşterisi" olan örgüt liderlerinin de erken safhada konuya dâhil olmasını gerektirir. Müşterilerin başlangıçta gürültünün yaygınlığı konusunda şüpheci olmasını bekleyebilirsiniz. Başlangıçtaki bu şüphecilik, beraberinde açık görüşlü bir tutum, denetimin sonuçlarıyla ilgili merak ve danışmanın kötümser beklentilerinin teyit edilmesi hâlinde durumun iyileştirilmesi için çaba harcamaya isteklilik getiriyorsa, aslında bir avantajdır.

Yargıçlar. Müşteriler denetlenecek bir ya da daha fazla birim belirler. Seçilen birim önemli sayıda "yargıç" yani şirket adına benzer yargı ve kararlar üretecek profesyoneller içermelidir. Yargıçlar aralarında yer değiştirebilmelidir; örneğin bir kişi bir vakayla ilgilenmeye müsait değilse, yerine bir başkası geçebilmeli ve benzer bir yargıya varması beklenmelidir. Bu kitabın sunduğu örnekler, federal yargıların cezai hüküm kararları,

bir sigorta şirketindeki risk primleri ve hasar rezervleriydi. Bir gürültü denetiminde, (1) yazılı bilgi temelinde tamamlanabilecek, (2) sayısal olarak (örneğin dolar, olasılık ya da puan olarak) ifade edilebilecek bir yargı görevi seçmek en iyisi olacaktır.

Proje müdürü. İdari personelden üst düzey bir yönetici, proje müdürü olarak atanmalıdır. Bu görev için belirli bir profesyonel uzmanlık şart değildir. Ancak örgütte yüksek bir pozisyon idari engelleri aşmak açısından pratik önem taşır ve ayrıca şirketin projeye verdiği önemi gösterir. Proje müdürünün görevi nihai raporun hazırlanması ve sonuçların şirketin liderlerine sunulması dâhil, projenin bütün safhalarını kolaylaştıracak idari desteği sağlamak olacaktır.

Vaka Materyallerinin Kurulması

Proje ekibinin parçası olan konu uzmanları birimin görevi konusunda kabul gören bir uzmanlığa sahip olmalıdır. (Örneğin riskler için prim belirlemek ya da olası yatırımların potansiyelini değerlendirmek gibi.) Denetimde kullanılacak vakaların geliştirilmesinden onlar sorumludurlar. Profesyonellerin işleri yaparken vardıkları yargıların inandırıcı bir simülasyonunu tasarlamak hassas bir iştir; özellikle ortaya ciddi sorunların çıkması hâlinde, çalışmanın nasıl titiz bir incelemeye maruz kalacağı düşünüldüğünde. Ekip şu soruyu düşünmeli: Simülasyonumuzun sonuçları yüksek bir gürültü seviyesini işaret ederse, şirketteki insanlar birimin gerçek yargılarında gürültü olduğunu kabul eder mi? Gürültü denetimi ancak cevap açıkça evetse yapılmaya değer.

Olumlu bir tepki almanın birden fazla yolu var. Bölüm 1'de tarif edilen cezai hüküm konulu gürültü denetimi, her vakayı, alakalı özelliklerin kısa bir şematik listesiyle özetledi ve on altı vaka için doksan dakikada değerlendirme aldı. Bölüm 2'de anlatı-

lan sigorta şirketindeki gürültü denetimi ise, karmaşık vakaların detaylı ve gerçekçi özetlerini kullandı. Her iki durumda da yüksek gürültü bulgularına ulaşılması, basitleştirilmiş vakalarda çok fazla anlaşmazlığa rastlanıyorsa gerçek vakalarda gürültünün çok daha kötü olacağı argümanı sayesinde kabul edilebilir kanıt sağladı.

Her vaka için, her yargıcın o vakada bir yargıya varırken nasıl akış yürüttüğü konusunda daha derin görüş sağlayacak bir soru formu hazırlanmalı. Soru formu sadece bütün vakaların tamamlanmasından sonra kullanılmalı. Form şunları içermeli:

Katılımcıyı cevabına yönlendiren kilit faktörlerle ilgili açık sorular.

Vakanın olgularının, katılımcıların önemlerini puanlamasına olanak sağlayan listesi.

Vakanın ait olduğu kategoriye "dış bakışı" gerektiren sorular. Örneğin: Eğer vaka dolar olarak değerlendirme gerektiriyorsa katılımcılar, vakanın, aynı kategorideki vakalar için bütün değerlerle karşılaştırıldığında ortalamanın ne kadar altında ya da üstünde olduğu konusunda bir tahmin sağlamalılar.

Üst Düzey Yöneticilerle Ön Toplantı

Denetimde kullanılacak vaka materyalleri bir araya toplanınca, proje ekibinin denetimi şirketin liderlerine sunacağı bir toplantı planlanmalıdır. O toplantıdaki tartışma, çalışmanın kabul edilemez bir sistem gürültüsünün bulunması dâhil, olası sonuçlarını göz önünde bulundurmalıdır. Toplantının amacı, planlanan çalışmaya itirazları dinlemek ve liderlerden sonuçları, –her ne olurlarsa olsunlar– kabul etme taahhüdünü almaktır. Bu tür bir taahhüt sağlanmadan bir sonraki aşamaya geçmenin bir anlamı olmaz.

Eğer ciddi itirazlar doğarsa, proje ekibinden vaka materyallerini iyileştirmeleri ve yeniden denemeleri talep edilebilir.

Üst düzey yöneticiler gürültü denetiminin tasarımını kabul edince, proje ekibi onlardan çalışmanın sonucuyla ilgili beklentilerini belirtmelerini istemelidir. Şu tür sorular tartışılmalıdır:

> "Her vaka için rastgele seçilmiş cevap çiftleri arasında hangi düzeyde anlaşmazlık bekliyorsunuz?"
>
> "İş alanı bakış açısıyla, kabul edilebilir en yüksek anlaşmazlık seviyesi nedir?"
>
> "Her iki istikamette yanlış bir değerlendirmeye ulaşılmasının (çok yüksek ya da çok düşük) tahminî maliyeti belirli bir miktara göre nedir (Mesela %15)?"

Bu soruların cevapları denetimin gerçek sonuçları geldiğinde hatırlanmaları ve inanılmaları için belgelenmelidir.

Çalışmanın Yürütülmesi

Denetlenen birimin yöneticileri en başından itibaren, birimlerinin özel bir çalışma için seçildiği konusunda genel anlamda bilgilendirilmelidir. Ancak, proje tarif edilirken *gürültü denetimi* ifadesinin kullanılmaması çok önemlidir. *Gürültü* ve *gürültülü* kelimelerinden, özellikle insanları tanımlarken, kaçınılmalıdır. Bunun yerine *karar alma çalışması* gibi nötr bir terim tercih edilmelidir.

Birimin yöneticileri, proje müdürü ve proje ekibinin katılımıyla, derhâl veri toplanmasından ve katılımcıları görev konusunda bilgilendirmekten sorumlu olacaklardır. Egzersizin niyeti katılımcılara, *"Örgüt karar alıcıların çıkarımlarına nasıl ulaştıklarını öğrenmekle ilgileniyor."* gibi genel bir ifadeyle anlatılmalıdır.

Çalışmaya katılan profesyonellere bireysel cevaplarının proje ekibi dâhil örgütten kimse tarafından bilinmeyeceği güvencesini

vermek esastır. Gerekirse, veriyi anonimleştirmek için dışarıdan bir firma tutulabilir. Ayrıca, sadece örgüt adına yargı görevlerini yerine getiren bütün birimlerin bir temsilcisi olarak seçilen birim için özel bir sonucunun olmayacağını vurgulamakta da fayda olacaktır. Sonuçların inandırıcılığını sağlama almak için, birimdeki bütün kalifiye profesyonel çalışmaya katılmalıdır. Egzersize yarım iş günü ayrılması katılımcıları önemine ikna etmeye yardımcı olacaktır.

Bütün katılımcılar egzersizi aynı zamanda tamamlamalıdır ama fiziksel olarak birbirlerinden ayrı tutulmalı ve çalışma süresince iletişim kurmamaları rica edilmelidir. Proje ekibi çalışma sırasında sorulara cevap vermek için erişilebilir olmalıdır.

Analizler ve Çıkarımlar

Her katılımcının değerlendirdiği birden fazla vakanın istatistiki analizinden, toplam gürültü miktarının ve ögeleri yatay gürültü ve örüntü gürültüsünün ölçümü dâhil, proje ekibi sorumlu olacaktır. Vaka materyalleri izin verirse, ayrıca cevaplardaki istatistiksel ön yargıları da saptayacaktır. Proje ekibi yine eşit derecede önemli bir görev olarak, katılımcıların akıl yürütme şekillerini izah ettikleri ve kararlarını en çok etkileyen olguları belirttikleri anket cevaplarını inceleyerek, yargılardaki değişkenliğin kaynaklarını anlamaya da çalışmalıdır. Ağırlıklı olarak dağılımın iki ucunda kalan cevaplara odaklanarak veride örüntüler aramalıdır. Çalışanların eğitiminde, örgütün prosedürlerinde ve çalışanlarına sağladığı bilgilerde olası eksikliklerin işaretlerini taramalıdır.

Danışman ve iç proje ekibi, birimde varılan yargı ve alınan kararları iyileştirmek için, karar hijyeni ve ön yargıyı giderme prensiplerini uygulayan araç ve prosedürleri geliştirmek için birlikte çalışmalıdır. Sürecin bu adımının birkaç aya yayılması olasıdır.

Paralel olarak danışman ve profesyonel ekip, proje hakkında bir rapor hazırlayarak örgütün liderine sunacaktır.

Bu noktada, örgüt birimlerinden birinde örnek bir gürültü denetimi yapmış olacak. Çaba başarılı kabul edilirse, yönetim ekibi örgütte üretilen yargı ve kararların kalitesini değerlendirmek ve iyileştirmek için daha geniş kapsamlı bir girişimde bulunabilir.

EK B

Karar Gözlemcisi Bir Kontrol Listesi

Bu ek bir karar gözlemcisi tarafından kullanılabilecek genel bir kontrol listesi örneği içermektedir (Bakınız Bölüm 19). Burada sunulan kontrol listesi, kabaca, önemli bir kararın yolunu açan tartışmanın kronolojik dizilimini de takip etmektedir.

Listedeki her maddeyi izleyen soru önerisi ek açıklamalar getirir. Karar gözlemcileri karar sürecini gözlemlerken kendilerine bu soruları sormalıdırlar.

Liste olduğu gibi kullanılmak üzere hazırlanmadı. Daha çok kendilerine ait, duruma özel bir ön yargı kontrol listesi tasarlayacak karar gözlemcilerine ilham ve başlangıç noktası sağlamasını umuyoruz.

Ön Yargı Gözlem Kontrol Listesi

1. YARGIYA YAKLAŞIM

1a. İkme

— "Grubun kanıt seçimi ve tartışmalarının odağı, görevlendirildikleri zor sorunun yerini daha kolay bir tanesinin aldığına işaret ediyor mu?"

— "Grup önemli bir faktörü ihmal etti mi? (ya da alakasız bir faktöre ağırlık verir gibi göründü mü?)"

1b. İç bakış

— "Grup müzakeresinin bir kısmı için dış görüş benimsedi mi ve mutlak yargı yerine karşılaştırmalı yargı uygulamak için ciddi girişimde bulundu mu?"

1c. Görüşlerin çeşitliliği

— "Grup üyelerinin hatalarının birbiriyle bağıntılı olmasına yol açabilecek herhangi ön yargı paylaştıklarından şüphelenmek için bir neden var mı? Tam tersi şekilde, aklınıza bu grupta temsil edilmeyen alakalı bir bakış açısı ya da uzmanlık geliyor mu?"

2. PEŞİN HÜKÜMLER VE VAKTİNDEN ÖNCE KAPATMA

2a. Başlangıçtaki peşin hükümler

— "Karar alıcılar(dan herhangi biri) bir çıkarımdan diğerinden daha fazla kazanç edebilecek gibi görünüyor mu?"

— "Aralarında bir çıkarıma zaten bağlanmış olan var mıydı? Peşin hükümden şüphelenmek için bir neden var mı?"

— "Muhalifler görüş belirtti mi?"

— "Kaybetmesi muhtemel bir eylem planına bağlılığın artması riski var mı?"

2b. Henüz olgunlaşmadan vaktinden önce kapatma; aşırı tutarlılık

— "Erken tartışılan konuların seçiminde tesadüfi ön yargı var mıydı?"

— "Alternatifler tam olarak değerlendirildi mi ve onları destekleyecek kanıtlar aktif olarak arandı mı?"

— "Rahatsız edici veri ve görüşler bastırıldı ya da yok sayıldı mı?"

3. BİLGİYİ İŞLEME

3a. Erişilebilirlik ve öne çıkma

— "Katılımcılar bir olayın alakasını, diyagnostik olmamasına rağmen, yakın zamanda gerçekleşmesi, dramatik kalitesi ya da kişisel alakası açısından abartıyor mu?"

3b. Bilginin kalitesine dikkatsizlik

— "Yargı ağırlıklı olarak anekdotlara, hikâyelere ya da benzetmelere mi dayanıyordu? Veriler onları teyit etti mi?"

3c. Çapa atma

— "Belirsiz bir isabetliliğe ya da alakaya sahip sayılar son yargıda önemli bir rol oynadı mı?"

3d. Regresif olmayan öngörü

— "Katılımcılar nonregresif dışdeğer tahminleri, tahminler, öngörüler yaptılar mı?"

4. KARAR

4a. Planlama yanılgısı

— "Tahminler kullanılırken insanlar kaynaklarını ve geçerliliklerini sorguladılar mı? Tahminlere meydan okumak için dış görüş kullanıldı mı?"

— "Belirsiz sayılar için güven aralıkları kullanıldı mı? Yeterince geniş miydiler?"

4b. Kayıptan kaçınma

— "Karar alıcıların risk iştahı örgütünkine uyuyor mu? Karar ekibi aşırı temkinli mi?"

4c. Mevcut ön yargı

— "Hesaplamalar (kullanılan indirim oranı dâhil), örgütün kısa ve uzun vadeli önceliklerinin dengesini yansıtıyor mu?"

EK C

Tahminleri Düzeltmek

Eşleşen öngörüler, sezgisel eşleştirme sürecine güvenimizden kaynaklanan hatalardır (bakınız Bölüm 14). Bir tahminde bulunmak için elimizdeki bilgiye bel bağladığımız ve o bilgi sonuç konusunda kusursuz şekilde (ya da fazlasıyla) kestirimciymiş gibi davrandığımız zaman, eşleşen öngörülerde bulunuruz.

"Dört yaşındayken akıcı şekilde okuyabilen" Julie örneğini hatırlayın. Soru, not ortalamasının kaç olduğuydu. Julie'nin üniversite not ortalaması için 3,8 tahmininde bulunduysanız, sezgisel olarak dört yaşındaki Julie'nin okuma konusunda kendi yaş grubunun en yüksek %10'luk diliminde (%3 ya da %5 değil) olduğu yargısına varmışsınızdır. Daha sonra, üstü örtülü bir şekilde, Julie'nin not ortalaması açısından da sınıfının en üst %10'luk diliminde olduğunu varsaydınız. Bu da 3,7 ya da 3,8'lik bir not ortalamasına karşılık geliyor. Bu cevapların popülerliği bundan kaynaklanıyor.

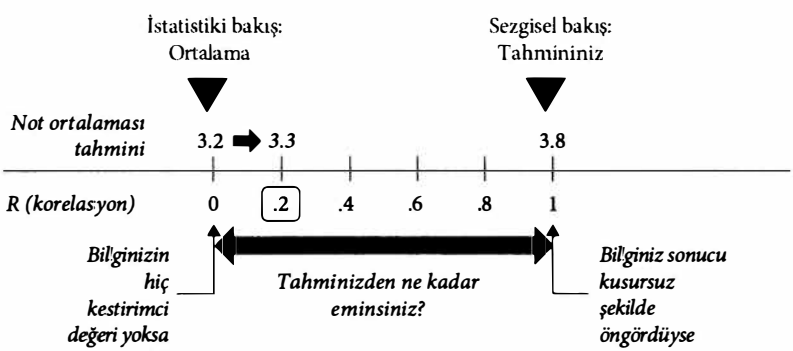
Bu akıl yürütme şeklini istatistiki açıdan hatalı kılan, Julie hakkındaki erişilebilir bilginin teşhissel değerini fazlasıyla abartmasıdır. Dört yaşındayken, yaşıtlarından ileri bir gelişim sergileyen her çocuk akademik açıdan üstün başarılı olmaz (ve çok şükür,

okuma konusunda zorlanan her çocuk da sonsuza dek sınıfının en altında kalmaz).

Aslında, dikkat çekici performans genellikle daha az dikkat çekici bir hâl alır. Tam tersine, çok kötü performans da gelişme kaydeder. Bu gözlem için sosyal, psikolojik ve hatta politik nedenleri hayal etmek hiç zor değil ama bizden nedenler istenmiyor. Bu, tamamen istatistiksel bir fenomendir. Bir yöndeki uç gözlemler, daha az uç olacaklardır çünkü geçmiş performans gelecek performansla kusursuz bir şekilde bağıntılı değildir. Bu eğilime *ortalamaya regresyon* adı verilir. (Bu noktayı hesaba katmayan eşleşen öngörüler için teknik terimin *regresif olmayan* olması da bu yüzdendir.)

Sayısal olarak ifade edecek olursak, Julie hakkında vardığınız yargı, eğer okuma yaşı not ortalaması için kusursuz bir kestirici olsaydı (yani iki faktör arasında korelasyon 1 olsaydı) doğru olurdu.

Daha isabetli olması olası bir yargıda bulunmanın istatistiksel bir yolu var. Bu, sezgiye dayanmayan ve istatistik eğitimi almış insanlar için bile bulması zor bir yol. Prosedürü aşağıda bulabilirsiniz. Şekil 19 bu yolu Julie örneğiyle anlatıyor.



ŞEKİL 19: *Sezgisel bir öngörüyü ortalamaya regresyon için ayarlamak*

1. *Sezgisel tahminde bulunun.*

Julie'yle ya da bilgi sahibi olduğunuz herhangi bir vakayla ilgili sezginiz değersiz değildir. Hızlı, Sistem 1 düşünmeniz elinizdeki bilgiyi kolayca tahmin ölçeğinize yerleştirir ve Julie için bir not ortalaması puanı üretir. Bu tahmin, elinizdeki bilgi kusursuz bir şekilde kestirimci olsa yapacağınız tahmindir. Not edin.

2. *Ortalamayı arayın.*

Şimdi geri çekilin ve Julie hakkında bildiklerinizi bir an için unutun. *Hakkında hiçbir şey bilmeseydiniz*, Julie'nin ortalaması için ne derdiniz? Cevap, elbette, çok açık: Hiçbir bilgi olmadığında Julie'nin not ortalaması için en iyi tahmininiz mezun olduğu sınıfın ortalama notu olmalı. (Muhtemelen 3,2 civarı olacaktır.)

Julie'ye bu şekilde bakmak, yukarıda bahsettiğimiz daha geniş bir prensibin (*dış bakış*) uygulanmasıdır. Dış bakış takındığımızda, değerlendirdiğimiz vakayı bir sınıfın örneği olarak ve sınıfı da istatistiksel olarak düşünürüz. Örneğin, hatırlarsanız, Gambardi sorununda dış bakış takınmak bizi yeni bir CEO için başarı baz oranının kaç olduğunu sormaya yönlendirmişti. (Bakınız Bölüm 4)

3. *Elinizdeki bilginin teşhissel değerini tahmin edin.*

Bu, kendinize, "Sahip olduğum bilginin kestirimci değeri nedir?" sorusunu sorduğunuz zor bir adım. Bu sorunun neden önemli olduğu, artık net olmalı. Julie hakkında tek bildiğiniz ayakkabı numarası olsaydı, bu bilgiye sıfır değer vermeniz ve ortalama not tahminine bağlı kalmanız tamamen doğru olurdu. Ancak öte yandan, Julie'nin her ders için aldığı notların bir listesine sahip olsaydınız, bu bilgi not ortalaması için kusursuz şekilde kestirimci olurdu (çünkü notların ortalamasını almanız yeterli olurdu). Bu iki uç nokta arasında grinin pek çok tonu vardır. Elinizde Julie'nin lisedeki olağanüstü entelektüel başarıları hakkında veri olsaydı,

bu bilgi teşhis konusunda okuma yaşından daha fazla, üniversite notlarından daha az değerli olurdu.

Burada göreviniz elinizdeki verinin teşhissel değerini, tahmin ettiğiniz sonuçla bir korelasyon olarak ifade edilecek bir sayısal değere dönüştürmek. Nadir vakalar dışında, bu sayı bir zarfın arka yüzünde yapılmış bir tahminden fazlası olamaz.

Makul bir tahmin için, Bölüm 12'de listelediğimiz bazı örnekleri hatırlayın. Sosyal bilimlerde 0,50'nin üstüne çıkan tahmin çok azdır. Anlamlı olarak gördüğümüz pek çok korelasyon 0,20 civarındadır. Julie'nin örneğinde, 0,20'lik bir korelasyon muhtemelen üst sınır olur.

4. Dış bakıştan, sezgisel yargınızın istikametinde, sahip olduğunuz bilginin teşhissel değerini yansıtacak düzeyde bir ayarlama yapın.

Son adım ürettiğiniz üç sayının aritmetik kombinasyonu: ortalamadan, sezgisel tahminize doğru, tahmin ettiğiniz korelasyonla orantılı olarak ayarlama yapmak.

Bu adım az önce yaptığımız bir gözlemi genişletiyor: Korelasyon 0 olsaydı, ortalamaya bağlı kalırdınız. 1 olsaydı, ortalamayı yok sayar ve mutlu bir şekilde eşleşen bir öngörüde bulunurdunuz. Öyleyse, Julie örneğinde, yapabileceğiniz en iyi not ortalaması tahmini, sınıfının ortalamasından, okuma yaşının size düşündürdüğü sezgisel tahmine doğru, en çok %20 uzaklaşan bir sayı olacak. Bu hesaplama sizi yaklaşık 3,3'lük bir sonuca ulaştırır.

Julie'nin örneğini kullandık ama bu yöntem bu kitapta ele aldığımız yargı sorunlarının çoğuna rahatça uygulanabilir. Örneğin, yeni bir satış temsilcisi almaya çalışan ve kesinlikle öne çıkan bir adayla mülakat yapmış bir satıştan sorumlu başkan yardımcısı düşünün. Bu güçlü izlenime dayanarak, başkan yardımcısı adayın işteki ilk yılında 1 milyon dolarlık bir satış getireceğini tahmin ediyor. Başkan yardımcısı bu tahmini nasıl regresif hâle

getirebilir? Hesaplama, mülakatın teşhissel değerine bağlı. Bir iş mülakatı bu örnekte işteki başarıyı ne kadar iyi öngörebilir? Gözden geçirdiğimiz kanıtlara dayanarak, 0,40'lık bir korelasyon çok cömert bir tahmin olur. Buna göre, işe yeni alınan adayın ilk yılki başarısının regresif tahmini en çok 500K $ + (1 milyon $ – 500K $) x 0,40 = 700K $ olur.

Bu süreç hiç de sezgisel değildir. Örneklerin gösterdiği gibi, düzeltilmiş tahminler her zaman sezgisel olanlardan daha tutucu olacaktır. Hiçbir zaman sezgisel tahminler kadar abartılı olamazlar, aksine, genellikle ortalamaya *çok daha* yakın olurlar. Tahminlerinizi düzeltirseniz, on kez Grand Slam Şampiyonu unvanı kazanmış bir tenis şampiyonunun 10 kez daha kazanacağı tahmininde bulunmazsınız. Tıpkı 1 milyar dolar değerindeki son derece başarılı bir yeni şirketin bunun yüzlerce katı büyüklüğünde bir deve dönüşeceği tahmininde bulunmayacağınız gibi. Düzeltilmiş tahminler uç değerde bahislerde bulunmaz.

Bu, geriye dönüp bakıldığında, düzeltilmiş tahminlerin kaçınılmaz olarak görülebilir hatalara neden olacağı anlamına gelir. Ancak tahmin geriye bakılarak yapılmaz. Uç değerlerin tanım gereği çok nadir olduğunu hatırlamalısınız. Tam tersi hata çok daha sıktır; uç değerlerin uç değer kalacağını öngördüğümüzde, ortalamaya regresyon nedeniyle, genellikle öyle kalmazlar. Bu nedenle, amaç isabetliliği en üst seviyeye çekmek olduğunda (yani OKH'yi en aza indirmek olduğunda) düzeltilmiş tahminler sezgisel, eşleşen öngörülerden üstün olacaktır.

Teşekkürler

Teşekkür edeceğimiz çok insan var. Linnea Gandhi bizim kurmay başkanlığımızı üstlenerek yabana atılmayacak miktarda rehberlik ve yardım sundu, organize olmamızı, gülümsememizi ve gülmemizi sağlayarak, kısaca şovun sürmesine neden oldu. Bunların yanı sıra metin üzerinde sayısız değerli öneride bulundu. Onsuz başaramazdık. Dan Lovallo da bu kitabın tohumunu atan makalelerden birinin eş yazarı olarak önemli bir rol üstlendi. Ajansımız John Brockman her aşamada hevesli, umutlu, keskin ve bilgeydi. Ona minnettarız. Baş editörümüz ve rehberimiz Tracy Behar bu kitabın irili ufaklı pek çok yönden büyük görünmesini sağladı. Arabella Pike ve Ian Straus da süper editoryal önerilerde bulundular.

Taslak bölümleri (bazen tam metnin bir müsveddesini) okuyup yorum yapan Oren Bar-Gill, Maya Bar-Hillel, Max Bazerman, Tom Blaser, David Budescu, Jeremy Clifton, Anselm Dannecker, Vera Delaney, Itiel Dror, Angela Duckworth, Annie Duke, Dan Gilbert, Adam Grant, Anupam Jena, Louis Kaplow, Gary Klein, Jon Kleinberg, Nathan Kuncel, Kelly Leonard, Daniel Levin, Sara McLanahan, Barbara Mellers, Josh Miller, Sendhil Mullainathan, Scott Page, Eric Posner, Lucia Reisch, Matthew Salganik, Eldar Shafir, Tali Sharot,

Philip Tetlock, Richard Thaler, Barbara Tversky, Peter Ubel, Crystal Wang, Duncan Watts ve Caroline Webb'e özel teşekkürlerimizi sunuyoruz. Cömertlikleri ve yardımları için minnettarız.

Pek çok büyük araştırmacının tavsiyelerinden faydalanma şansına sahip olduk. Julian Parris pek çok istatistiksel meselede paha biçilmez bir yardım sağladı. Yapay öğrenmenin başarılarını konu alan bölümlerimiz Sendhil Mullainathan, Jon Kleinberg, Jens Ludwig, Gregory Stoddard ve Hye Chang olmadan mümkün olmazdı. Yargının tutarlılığıyla ilgili tartışmamız Alex Todorov ve Princeton'lı çalışma arkadaşları Joel Martinez, Brandon Labbree ve Stefan Uddenberg kadar Scott Highhouse ve Alison Broadfoot'a da çok şey borçlu. Bu müthiş araştırma ekipleri sadece içgörülerini büyük bir cömertlikle paylaşmakla kalmadılar, üstüne bizim için bazı özel analizler yapma inceliğini de sergilediler. Elbette her tür yanlış anlaşılma ya da hata bizim sorumluluğumuzdur. Ek olarak, Laszlo Bock, Bo Cowgill, Jason Dana, Dan Goldstein, Harold Goldstein, Brian Hoffman, Alan Krueger, Michael Mauboussin, Emily Putnam-Horstein, Charles Scherbaum, Anne-Laure Sellier ve Yuichi Shoda'ya uzmanlıklarını paylaştıkları için çok teşekkür ediyoruz.

Ayrıca senelerdir destek aldığımız hatırı sayılır bir araştırmacı ordusuna da minnettarız: Shreya Bhardwaj, Josie Fisher, Rohit Goyal, Nicole Grabel, Andrew Heinrich, Meghann Johnson, Sophie Mehta, Eli Nachmany, William Ryan, Evelyn Shu, Matt Summers ve Noam Ziv-Crispel. Buradaki tartışmaların pek çoğu bizim yeterli uzmanlığa sahip olmadığımız somut alanları içeriyor ve onların kusursuz çalışmaları sayesinde kitap aksi takdirde olacağından daha az ön yargı ve daha az gürültü içeriyor.

Son olarak, üç yazar ve iki kıtaya ayrılmış bir takım olarak iş birliği yapmak en iyi zamanlarda bile çok zorken, 2020'nin o iyi zamanlardan olmadığını ortada. Dropbox ve Zoom'un teknolojik sihri olmadan bu kitabı asla bitiremezdik. Bu harika ürünlerin arkasındaki insanlara minnettarız.

Notlar

GİRİŞ

10. *Hedefler farkı görselleştiriyor*: İsviçreli matematikçi Daniel Bernouilli 1778'de tahmin sorunları üstüne bir denemesinde aynı benzetmeyi silahlar yerine ok ve yay kullanarak yaptı. Bernouilli, "The Most Probable Choice Between Several Discrepant Observations and the Formation Therefrom of the Most Likely Induction", *Biometrika 48*, no. 1-2 (Haziran, 1961): s. 3-18; https://doi.org/10.1093/biomet/48.1-2.3.

12. *Çocuk velayeti kararları*: Joseph J. Doyle Jr., "Child Protection and Child Outcomes: Measuring the Effects of Foster Care", *American Economic Review* 95, no. 5 (Aralık 2007): s. 1583-1610.

13. *Aynı yazılımcılara*: Stein Grimstad ve Magne Jørgensen, "Inconsistency of Expert Judgment-Based Estimate of Software Development Effort", *Journal of Systems and Software* 80, no. 11 (2007): s. 1770-1777.

13. *Sığınma kararları*: Andrew I. Schoenholtz, Jaya Ramji-Nogales ve Philip G. Schrag, "Refugee Roulette: Disparities in Asylum Adjudication", *Stanford Law Review* 60, no. 2 (2007).

14. *Patent verme kararları*: Mark A. Lemley ve Bhaven Sampat, "Examiner Characteristics and Patent Office Outcomes", *Review of Economics and Statitics* 94, no. 3 (2012): s. 817-827. Bakınız ayrıca Iain Cockburn, Samuel Kortum ve Scott Stern, "Are All Patent Examiners Equal? The Impact of Examiner Characteristics", makale taslağı 8980, Haziran 2002; www.nber.org/papers/w8980 ve Michael D. Frakes ve Melissa F. Wasserman, "Is

the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents? Evidence from Microlevel Application Data", *Review of Economics and Statistics* 99, no. 3 (Temmuz 2017): s. 550-563.

BÖLÜM 1

22. *Motivasyonunu şu şekilde anlatıyordu*: Marvin Frankel, *Criminal Sentences: Law Without Order*, 25 The Institute for Scientific Information Current Contents / Social & Behavioral Sciences: This Week's Citation Classic 14, 2A-6 (23 Haziran 1986); http://www.garfield.library.upenn.edu/classics1986/A1986C697400001.pdf.
23. *"neredeyse tamamen kontrolsüz"*: Marvin Frankel, *Criminal Sentences: Law Without Order*, New York: Hill and Wang, 1973: s. 5.
23. *"her gün işlenen keyfî zulümlere"*: Frankel, *Criminal Sentences*, s. 103.
23. *"insanlara değil, hukuka dayalı bir hükûmet"*: Frankel, s. 5.
23. *"nevi şahsına münhasır fermanlar"*: Frankel, s. 11.
23. *"bir tür sayısal ya da benzer nesnel puanlama"*: Frankel, s. 114.
23. *"bilgisayarların... yardımcı olarak"*: Frankel, s. 115.
23. *hüküm verme konusunda bir komisyon*: Frankel, s. 119.
24. *"uzlaşmanın olmamasının kural olduğu"*: Anthony Partridge ve William B. Eldridge, *The Second Circuit Sentence Study: A Report to the Judges of the Second Circuit August 1974*, Washington, DC: Federal Judicial Center, Ağustos 1974: s. 9.
24. *"şaşırtıcı"*: ABD Senatosu, "Comprehensive Crime Control Act of 1983: Report of the Committee on the Judiciary, United States Senate, on S. 1762. Together with Additional and Minority Views", Washington, DC: US Government, Printing Office, 1983, Rapor No: 98-225.
24. *Bir eroin satıcısı*: Anthony Partridge ve Eldridge, *Second Circuit Sentence Study*, A-11.
24. *bir banka soyguncusu*: Partridge ve Eldridge, *Second Circuit Sentence Study*, A-9.
24. *bir şantaj vakası*: Partridge ve Eldridge, *Second Circuit Sentence Study*, A-5 - A-7.
24. *kırk yedi yargıçla bir çalışma*: William Austin ve Thomas A. Williams III, "A Survey of Judges' Responses to Simulated Legal Cases: Research Note on Sentencing Disparity", *Journal of Criminal Law & Criminology* 68 (1977): s. 306.

24. *çok daha geniş ölçekli bir çalışma*: John Bartolomeo ve diğerleri, "Sentence Decisionmaking: The Logic of Sentence Decisions and the Extent and Sources of Sentence Disparity", *Journal of Criminal Law and Criminology* 72 no. 2 (1981). (Tartışmanın tamamı için bakınız Bölüm 6) Ayrıca bakınız Senato Raporu, 44.
25. *Yargıçlar açken*: Shai Danziger, Jonathan Levav ve Liora Avnaim-Pesso, "Extraneous Factors in Judicial Decisions", *Proceedings of the National Academy of Sciences of the United States of America* 108, no. 18 (2011): s. 6889-92.
25. *Binlerce çocuk mahkemesi kararı*: Özkan Eren ve Naci Mocan, "Emotional Judges and Unlucky Juveniles", *American Economic Journal: Applied Economics* 10, no. 3 (2018): s. 171-205.
25. *yenilmesi hâlinde... daha sert kararlar*: Daniel L. Chen ve Markus Loecher, "Mood and the Malleability of Moral Reasoning: The Impact of Irrelevant Factors on Judicial Decisions", *SSRN Electronic Journal* (21 Eylül 2019): s. 1-70; http://users.nber.org/dlchen/papers/Mood_and_the_Malleability_of_Moral_Reasoning.pdf.
26. *doğum günlerinde daha da müsamahalı*: Daniel L. Chen ve Arnaud Philippe, "Clash of Norms: Judicial Leniency on Defendant Birthdays", (2020): https://ssrn.com/abstract=3203624.
26. *Dışarıdaki hava sıcaklığı gibi alakasız şeyler*: Anthony Heyes ve Soodeh Saberian, "Temperature and Decisions: Evidence from 207.000 Court Cases", *American Economic Journal: Applied Economics* 11, no. 2 (2018): s. 238-265.
26. *"sınırsız takdir yetkisini"*: Senato Raporu, 38.
26. *"açıklanamayacak şekilde yaygın" bir hüküm uyumsuzluğundan*: Senato Raporu, 38.
27. *geçmiş pratiklerin kullanılmasını*: Jeffrey Rosen, "Breyer Restraint", *New Republic*, 11 Temmuz 1994, s. 19 ve 35'te yer alan Yargıç Breyer'den alıntı yapılmıştır.
27. *dışına çıkmaları... tasdiklenmesi gerekiyordu*: ABD Ceza Komisyonu, Kılavuz El Kitabı (2018); www.ussc.gov/sites/default/files/pdf/guidelines-manual /2018/GLMFull.pdf.
28. *"net hüküm çeşitliliklerini azaltmıştı"*: James M. Anderson, Jeffrey R. Kling ve Kate Stith, "Measuring Interjudge Sentencing Disparity: Before and After the Federal Sentencing Guidelines", *Journal of Law and Economics* 42, no. S1(Nisan 1999): s. 271-308.

28. *komisyonun kendisinden*: ABD Ceza Komisyonu, *The Federal Sentencing Guidelines: A Report on the Operation of the Guidelines System and Short-Term Impacts on Disparity in Sentencing, Use of Incarceration, and Prosecutorial Discretion and Plea Bargaining*, cilt 1 ve 2 , Washington, DC: ABD Ceza Komisyonu, 1991.
28. *Bir başka çalışmaya göre*: Anderson, Kling ve Stith, "Interjudge Sentencing Disparity".
28. *bağımsız bir çalışma*: Paul J. Hofer, Kevin R. Blackwell ve R. Barry Ruback, "The Effect of the Federal Sentencing Guidelines on Inter-Judge Sentencing Disparity", *Journal of Criminal Law and Criminology* 90 (1999): s. 239, 241.
28. "*körlüğe değil... ihtiyaç...*": Kate Stith ve José Cabranes, *Fear of Judging: Sentencing Guidelines in the Federal Courts*, Chicago: University of Chicago Press, 1998: s. 79.
29. *Yüksek Mahkeme'nin... kılavuzu lağvetmesine*: 543 U.S.220 (2005).
29. *%75'i, bu mecburi rejimin tavsiye niteliğinde kalmasını tercih ederken*: ABD Ceza Komisyonu "Results of Survey of United States District Judges, January 2010 through March 2010", (Haziran 2010) (soru 19, tablo 19); www.ussc.gov/sites/default/files/pdf /research- and- publications/ research- projects- and- surveys/surveys/20100608_Judge_Survey.pdf.
29. *bulgusunun... arttığı oldu*: Crystal Yang, "Have Interjudge Sentencing Disparities Increased in an Advisory Guidelines Regime? Evidence from Booker", *New York University Law Review* 89 (2014): s. 1268-1342, s. 1278 ve s. 1334.

BÖLÜM 2

35. *Gürültü denetimine hazırlanmak için*: Şirketin üst düzey yöneticileri, çalışanların her gün uğraştığı risk ve taleplerin benzeri olan temsilci vakaların detaylı tanımlarını oluşturdular. Mal ve Kaza Sigortası bölümündeki sigorta eksperleri için altı ve finansal riskleri hesaplamak için risk değerlendirme uzmanları için dört vaka hazırlandı. Çalışanlara her birinin iki ya da üç vakayı değerlendirmesi için, her zamanki iş yüklerinden yarım gün izin verildi. Bağımsız çalışmaları istendi ve çalışmanın amacının yargılarının değişkenliğini incelemek olduğu söylenmedi. Toplamda kırk sekiz risk değerlendirme uzmanından seksen altı ve altmış sekiz sigorta eksperi yüz on üç yargı aldık.

41. *Naif gerçekçilik*: Dale W. Griffin ve Lee Ross, "Subjective Construal, Social Inference, and Human Misunderstanding", *Advances in Experimental Social Psychology* 24 (1991): s. 319-359; Robert J. Robinson, Dacher Keltner, Andrew Ward ve Lee Ross, "Actual Versus Assumed Differences in Construal: 'Naive Realism' in Intergroup Perception and Conflict", *Journal of Personality and Social Psychology* 68 no. 3 (1995): s. 404 ve Lee Ross veAndrew Ward, "Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding", *Values and Knowledge* (1997).

KISIM 2

53. *değişkenliğin en yaygın ölçüsü*: Bir grup sayının standart sapması varyans denilen başka bir istatistiksel nicelikten türetilir. Varyansı hesaplamak için, önce ortalamadan sapmaların dağılımını elde etmeli sonra da bu sapmaların her birinin karesini almalıyız. Varyans bu sapmaların karelerinin ortalamasıdır ve standart sapma varyansın karekökü dür.

BÖLÜM 4

56. *Şarap yarışmalarında jüriler*: R. T. Hodgson, "An Examination of Judge Reliability at a Major U.S. Wine Competition", *Journal of Wine Economics* 3, no. 2 (2008): s. 105-103.
64. *alışverişler değerlendirmeci yargılarla çözüme kavuşturulur:* Bazı karar alma öğrencileri kararları seçenekler arasındaki tercihler olarak görür ve nicel yargıları olası seçeneklerin sürekliliğinin olduğu özel bir karar örneği olarak kabul eder. Bu bakış açısından, yargılar özel bir karar vakasıdır. Bizim buradaki yaklaşımımız farklı: Biz seçenekler arasında seçim yapmayı gerektiren kararları her seçenek için altta yatan değerlendirmeci bir yargıdan kaynaklanıyormuş gibi kabul ediyoruz. Yani kararları özel bir yargı vakası olarak görüyoruz.

BÖLÜM 5

72. *1795 yılında keşfedilen*: En küçük kareler yöntemi ilk kez Adrien-Marie Legendre tarafından 1805'te yayımlandı. Gauss kendisinin on yıl önce kullandığını iddia etti ve daha sonra bir hata teorisinin geliştirilmesiyle ve adını taşıyan çan eğrisiyle ilişkilendirdi. Öncelik konusu çok tartışıldı ve tarihçiler Gauss'un iddiasına inanmaya daha meyilliler. Stephen M. Stigler, "Gauss and the Invention of Least Squares", *Annals of Statistics*

9 [1981]: s. 465-474 ve Stephen M. Stigler, *The History of Statistics: The Measurement of Uncertainty Before 1900*, Cambridge, MA: Belknap Press of Harvard University Press, 1986.

76. *Basit cebir kullanılarak*: *Gürültü*yü hataların standart sapması olarak tanımladık, dolayısıyla gürültü karesi hataların varyansıdır. *Varyans*'ın tanımı "karelerin ortalaması eksi ortalamanın karesi"dir. Ortalama hata ön yargı olduğuna göre, "ortalamanın karesi" ön yargının karesidir. Bu nedenle: Gürültü^2 = OKH – Ön Yargı^2.

79. *bu açıdan sezgileri*: Berkeley J. Dietvorst ve Soaham Bharti, "People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error", *Psychological Science* 31, no. 10 (2020): s. 1302-1314.

BÖLÜM 6

83. *olağanüstü detaylı bir gürültü denetimi*: Kevin Clancy, John Bartolomeo, David Richardson ve Charles Wellford, "Sentence Decisionmaking: The Logic of Sentence Decisions and the Extent and Sources of Sentence Disparity", *Journal of Criminal Law and Criminology* 72, no. 2 (1981): s. 524-554 ve NSLAW, Inc. ve diğerleri, "Federal Sentencing: Towards a More Explicit Policy of Criminal Sanctions III-4" (1981).

84. *bir hüküm belirlemeleri istendi*: Hüküm her tür hapis süresi, gözetim süresi ve para cezası kombinasyonu olabilirdi. Basitleştirmek için, burada ağırlıklı olarak hükümlerin başlıca ögesine -hapis süresine- odaklanacağız ve diğer iki ögeyi kenara alacağız.

85. *bu değişkenlikten... sıklıkla..*: Çoklu vaka, çoklu yargıç ortamında, Bölüm 5'te bahsettiğimiz hata denkleminin genişletilmiş versiyonu bu değişkenliği yansıtan bir terim içerir. *Büyük bir ön yargıyı* bütün vakaların ortalama hatası olarak tanımlarsak ve bu hata vakadan vakaya aynı değilse, vaka ön yargıları arasında değişkenlik olur. Denklem: OKH = Büyük Ön Yargı^2 + Vaka Ön Yargılarının Değişkenliği + Sistem Gürültüsü^2 olur.

85. *Ortalama hapis cezası*: Bu bölümde bahsedilen sayılar orijinal çalışmadan aşağıdaki gibi türetildi.

Öncelikle yazarlar *suç ve suçlunun* ana etkisinin toplam varyansın %45'inden sorumlu olduğunu bildiriyorlar. (John Bartolomeo ve diğerleri, "Sentence Decisionmaking: The Logic of Sentence Decisions and the Extent and Sources of Sentence Disparity", *Journal of Criminal Law and Criminology* 72, no. 2 [1981], Tablo 6). Ancak burada biz daha geniş anlamda,

her vakanın, yargıçlara sunulan bütün özellikleri (suçlunun sabıkasının olup olmadığı suç işlenirken bir silahın kullanılıp kullanılmadığı gibi) kapsayan etkisiyle ilgileniyoruz. Bizim tanımımızla, bütün bu özellikler gürültü değil, *gerçek vaka varyansları*. Buna göre, her vakanın özellikleri arasındaki etkileşimleri vaka varyansında yeniden birleştirdik. (Bunlar toplam varyansın %11'ini oluşturuyor, bakınız Bartolomeo ve diğerleri, Tablo 10). Sonuç olarak vaka varyansının ve yargıç ana etkisinin paylarını sırasıyla %56 ve %21 olarak ve toplam varyanstaki etkileşimleri %23 olarak yeniden tanımladık. Dolayısıyla sistem gürültüsü toplam varyansın %44'ünü oluşturuyordu.

Adil hükümlerin varyansı Bartolomeo ve diğerleri, s. 89'dan hesaplanabilir; her vaka için ortalama hüküm sürelerini listeleyen tabloda varyans 15'tir. Eğer bu toplam varyansın %6'sı ise, o zaman toplam varyans 26,79'dur ve sistem gürültüsünün varyansı 11,79'dur. O varyansın karekökü temsilci bir vakanın standart sapması ya da 3,4 yıldır.

Yargıç ana etkisi ya da yatay gürültü toplam varyansın %21'idir. O varyansın karekökü yargıç yatay gürültüsüne atfedilebilecek standart sapma ya da 2,4 yıldır.

86. *3,4 yıldı*: Bu değer on altı vaka için hüküm sürelerinin varyanslarının ortalamasının karekökü dür. Bir önceki notta açıklandığı gibi hesapladık.

89. *basit ve eklemeli mantık*: Ekleme hipotezi bir yargıcın sertliğinin sabit bir hapis süresi eklediğini varsayar. Bu hipotezin doğru olması olası değildir: Bir yargıcın sertliğinin ortalama hüküm süresiyle orantılı bir miktar eklemesi daha olasıdır. Bu konu orijinal raporda ihmal edildiği için, önemini değerlendirme imkânı yok.

90. *yargıçlar arasında... örüntü farkları*: Bartolomeo ve diğerleri, "Sentence Decisionmaking", s. 23.

91. *neredeyse eşit şekilde*: Şu denklem geçerli: (Sistem Gürültüsü)^2 = (Yatay Gürültü)^2 + (Örüntü Gürültüsü)^2. Tablo sistem gürültüsünün 3,4 yıl ve yatay gürültünün 2,4 yıl olduğunu gösteriyor. Yani örüntü gürültüsü yaklaşık 2,4 yıl. Hesap bir çizim olarak gösteriliyor: Gerçek değerler yuvarlama hataları nedeniyle çok az farklıdır.

BÖLÜM 7

94. *Bütün zamanların en iyi*: Bakınız http://www.iweblists.com/sports/basketball/FreeThrow Percent_c.html, 27 Aralık 2020'de bakıldı.

94. *Shaquille O'Neal*: Bakınız https://www.basketball-reference.com/players/o/onealsh01.html, 27 Aralık 2020.
95. *şarap uzmanları*: *R.* T. Hodgson, "An Examination of Judge Reliability at a Major U.S. Wine Competition", *Journal of Wine Economics* 3, no. 2 (2008): s. 105-113
96. *yazılım danışmanları*: Stein Grimstad ve Magne Jørgensen, "Inconsistency of Expert Judgment-Based Estimates of Software Development Effort", *Journal of Systems and Software* 80, no. 11 (2007): s. 1770-1777.
97. *doğru cevap eğiliminde olma:* Robert. H. Ashton, "A Review and Analysis of Research on the Test-Retest Reliability of Professional Judgment", *Journal of Behavioral Decision Making* 294, no. 3 (2000) s. 277-294. Bu arada, yazar incelediği kırk bir çalışmadan birinin bile durum gürültüsünü değerlendirmek için tasarlanmadığını belirtiyor: "Her durumda, güvenilirliğin ölçümü diğer bazı araştırma amaçlarının bir yan ürünüydü". (Ashton, s. 279) Bu yorum durum gürültüsünü araştırmaya ilginin nispeten yakın zamanda başladığını düşündürüyor."
97. *doğru cevap*: Central Intelligence Agency, *The World Factbook*, Washington, DC: Central Intelligence Agency, 2020. Verilen sayı havadan ayırt edilebilen bütün havaalanı ve hava pistlerini içeriyor. Kalkış pistleri asfalt kaplı olabilir ya da olmayabilir ve kapalı ya da terk edilmiş tesisleri de içerebilir.
98. *Edward Vul ve Harold Pashler*: Edward Vul ve Harold Pashler, "Crowd Within: Probabilistic Representations Within Individuals".
98. *gerçeğe... daha yakındı*: James Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*, New York: Doubleday, 2004.
99. *daha az gürültülü*: Ortalaması alınan yargıların standart sapması (gürültü ölçümümüz) yargı sayısının karekökü ne orantılı olarak azalır.
99. *elde edebilirsiniz*: Vul ve Pashler, "Crowd Within", s. 646.
100. *Stefan Herzog ve Ralph Hertwig*: Stefan M. Herzog ve Ralph Hertwig, "Think Twice and Then: Combining or Choosing in Dialectical Bootstrapping?", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 40, no. 1 (2014): s. 218-232.
101. *verilen cevaplar*: Vul ve Pashler, "Measuring the Crowd Within", s. 647.
102. *Joseph Forgas*: Joseph P. Forgas, "Affective Influences on Interpersonal Behavior", *Psychological Inquiry* 13, no. 1 (2002): s. 1-28.
102. *bir gülümseme*: Forgas, "Affective Influences", s. 10.

102. *iyi bir ruh hâlinden... (pazarlıkçılar)*: A. Filipowicz, S. Barsade ve S. Melwani, "Understanding Emotional Transitions: The Interpersonal Consequences of Changing Emotions in Negotiations", *Journal of Personality and Social Psychology* 101, no. 3 (2011) s. 541-556.

103. *katılımcılar... kısa bir felsefi deneme okudular*: Joseph P. Forgas, "She Just Doesn't Look like a Philosopher..? Affective Influences on the Halo Effect in Impression Formation", *European Journal of Social Psychology* 41, no. 7 (2011): s. 812-817.

103. *sözde derin beyanlara:* Gordon Pennycook, James Allan Cheyne, Nathaniel Barr, Derek J. Koehler ve Jonathan A. Fugelsang, "On the Reception and Detection of Pseudo-Profound Bullshit", *Judgment and Decision Making* 10, no. 6 (2015): s. 549-563.

103. *Boktanlık Üstüne*: Harry Frankfurt, *On Bullshit*, Princeton, NJ: Princeton University Press, 2005.

103. *görüntüde etkileyici savlardan*: Pennycook ve diğerleri, "Pseudo-Profound Bullsihit", s. 549.

103. *genel olarak kolay kandırılmaya*: Joseph P. Forgas, "Happy Believers and Sad Skeptics? Affective Influences on Gullibility", *Current Directions in Psychological Science* 28, no. 3 (2019): s. 306-313.

104 *görgü tanıkları*: Joseph P. Forgas, "Mood Effects on Eyewitness Memory: Affective Influences on Susceptibility to Misinformation", *Journal of Experimental Social Psychology* 41, no.6 (2005): s. 574-588.

104. *üst geçit problemi*: Piercarlo Valdesolo ve David Desteno, "Manipulations of Emotional Context Shape Moral Judgment", *Psychological Science* 17, no. 6 (2006): s. 476-477.

105. *uzun bir günün sonunda opioid ilaçlar*: Hannah T. Neprash ve Michael L. Barnett, "Association of Primary Care Clinic Appointment Time with Opioid Prescribing", *JAMA Network Open* 2, no. 8 (2019); Lindsey M. Philpot, Bushra A. Khokhar, Daniel L. Roellinger, Priya Ramar ve Jon O. Ebbert, "Time of Day Is Associated with Opioid Prescribing for Low Back Pain in Primary Care", *Journal of General Internal Medicine* 33 (2018): s. 1828.

105. *antibiyotik*: Jeffrey A. Linder, Jason N. Doctor, Mark W. Friedberg, Harry Reyes Nieva, Caroline Birks, Daniella Meeker ve Craig R. Fox, "Time of Day and the Decision to Prescribe Antibiotics", *JAMA Internal Medicine* 174, no. 12 (2014): s. 2029-2031.

105. *grip aşısı*: Rebecca H. Kim, Susan C. Day, Dylan S. Small, Christopher K. Snider, Charles A. L. Rareshide ve Mitesh S. Patel, "Variations in Influenza Vaccination by Clinic Appointment Time and an Active Choice Intervention in the Electronic Health Record to Increase Influenza Vaccination", *JAMA Network Open* 1, no. 5 (2018): s. 1-10.

106. *Kötü hava durumu*: Daha iyi hafızayla ilgili yorum için bakınız Joseph P. Forgas, Liz Goldenberg ve Christian Unkelbach, "Can Bad Weather Improve Your Memory? An Unobtrusive Field Study of Natural Mood Effects on Real-Life Memory", *Journal of Experimental Social Psychology* 45, no. 1 (2008): s. 254-257. Güneş ışığıyla ilgili yorum için bakınız David Hirshleifer ve Tyler Shumway, "Good Day Sunshine: Stock Returns and the Weather", *Journal of Finance* 58, no. 3 (2003): s. 1009-1032.

106. *"Bulutlar İnekleri İyi Gösteriyor."*: "Clouds Make Nerds Look Good: Field Evidence of the Impact of Incidental Factors on Decision Making", *Journal of Behavioral Decision Making* 20, no. 2 (2007): s. 143-152.

106 *kumarbaz yanılgısı*: Daniel Chen ve diğerleri, "Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires", *Quarterly Journal of Economics* 131, no. 3 (2016): s. 1181-1242.

106. *sığınma hakkı tanıma*: Jaya Ramji-Nogales, Andrew I. Schoenholtz, and Philip Schrag, "Refugee Roulette: Disparities in Asylum Adjudication", *Stanford Law Review* 60, no. 2 (2007).

108. *hafıza performansı*: Michael J. Kahana ve diğerleri, "The Variability Puzzle in Human Memory", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 44, no. 12 (2018): s. 1857-1863.

BÖLÜM 8

112. *müzik indirme üzerine... bir çalışmadan...*: Matthew J. Salganik, Peter Sheridan Dodds ve Duncan J. Watts, "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market", *Science* 311 (2006): s. 854-856. Bakınız ayrıca Matthew Salganik ve Duncan Watts, "Leading the Herd Astray: An Experimental Study of Self-Fulfilling Prophecies in an Artificial Cultural Market", *Social Psychology Quarterly* 71 (2008): s. 338-355 ve Matthew Salganik ve Duncan Watts, "Web-Based Experiments for the Study of Collective Social Dynamics in Cultural Markets", *Topics in Cognitive Science* 1 (2009): s. 439-468.

113. *popülerlik kendi kendini pekiştirir*: Salganik ve Watts, "Leading the Herd Astray".

114. *daha pek çok alanda*: Michael Macy ve diğerleri, "Opinion Cascades and the Unpredictability of Partisan Polarization", *Science Advances* (2019): s. 1-8. Bakınız ayrıca Helen Margetts ve diğerleri, *Political Turbulence*, Princeton: Princeton University Press, 2015.

115. *sosyolog Michael Macy*: Michael Macy ve diğerleri, "Opinion Cascades".

115. *internet sitelerinde yer alan yorumları*: Lev Muchnik ve diğerleri, "Social Influence Bias: A Randomized Experiment", *Science* 341, no. 6146 (2013): s. 647-651.

117. *Araştırmalar tam olarak bu sorunu*: Jan Lorenz ve diğerleri, "How Social Influence Can Undermine the Wisdom of Crowd Effect", *Proceedings of the National Academy of Sciences* 108, no. 22 (2011): s. 9020-9025.

122. *karşılaştıran bir deney*: Daniel Kahneman, David Schkade ve Cass Sunstein, "Shared Outrage and Erratic Awards: Thee Psychology of Punitive Damages", *Journal of Risk and Uncertainty* 16 (1998): s. 49-86.

122. *beş yüzden fazla altı kişilik jüri*: David Schkade, Cass R. Sunstein ve Daniel Kahneman, "Deliberating about Dollars: The Severity Shift", *Columbia Law Review* 100 (2000): s. 1139-1175.

KISIM 3

128. *yüzdelik uyumdur*: Yüzdelik uyum (YU) uyum katsayısı olarak da bilinen *Kendall's W* ile yakından ilişkilidir.

128. *boy ve ayak ölçüsü*: Kanwal Kamboj ve diğerleri, "A Study on the Correlation Between Foot Length and Height of an Individual and to Derive Regression Formulae to Estimate the Height from Foot Length of an Individual", *International Journal of Research in Medical Sciences* 6, no.2 (2018): s. 528.

129. *Tablo 1... YU'yu sunuyor*: YU birleşik dağılımın iki değişkenli normal olduğu varsayımına göre hesaplanır. Tabloda gösterilen değerler bu varsayıma dayalı yaklaşıklardır. Bu tabloyu ürettiği için Julian Parris'e teşekkür ederiz.

BÖLÜM 9

132. *gerçek bir performans öngörüsü çalışmasına*: Martin C. Yu ve Nathan R. Kuncel, "Pushing the Limits for Judgmental Consistency: Comparing Random Weighting Schemes with Expert Judgments", *Personnel*

Assessment and Decisions 6 no. 2 (2020): s. 1-10. Uzmanların ulaştığı 0,15'lik korelasyon incelenen üç örneğin, toplamda 847 vakayı içeren ağırlıksız ortalamasıdır. Gerçek çalışma bu sadeleştirilmiş tanımdan birçok açıdan farklıdır.

133. *ağırlıklı ortalama*: Ağırlıklı bir ortalama oluşturmanın ön şartı bütün kestiricilerin karşılaştırılabilir birimlerle ölçülmesidir. Bu şart bütün puanlamaların 0 ila 10 ölçeğinde yapıldığı bizim örneğimizde sağlandı ama her zaman böyle olmaz. Örneğin performans kestiricileri bir mülakatçının 0 ila 10 ölçeğindeki değerlendirmesi, alakalı tecrübenin kaç yıl olduğu ve bir yeterlilik testinin puanı olabilir. Çoklu regresyon programı bütün kestiricileri birleştirmeden önce *standart puanlara* dönüştürür. Standart puan bir gözlemin bir nüfusun ortalamasına olan uzaklığını birim olarak standart sapmayla ölçer. Örneğin eğer yeterlilik testinin ortalaması 55'se ve standart sapma 8 ise, +1,5'luk bir standart puan testte 67 puana karşılık gelir. Her bireyin verisinin standardize edilmesi ortalamada ya da bireylerin yargılarının değişkenliğindeki herhangi bir hata ihtimalini ortadan kaldırır.

133. *büyük bir ağırlık alır*: Çoklu regresyonun önemli bir özelliği her kestirici için ideal ağırlığın diğer kestiricilere bağlı olmasıdır. Bir kestirici bir başkasıyla yüksek bir korelasyona sahipse, eşit derecede bir ağırlık almamalıdır: Bu bir çeşit "çifte sayım" olur.

133. *"... yük beygiri"*: Robin M. Hogarth ve Natalia Karelaia, "Heuristic and Linear Models of Judgment: Matching Rules and Environments", *Psychological Review* 114, no.3 (2007): s. 734.

133. *basit bir yapıyı:* Bu bağlamda yaygın olarak kullanılan bir araştırma çerçevesi, bu tartışmanın dayandığı *yargı lens modelidir.* Bakınız Kenneth R. Hammond, "Probabilistic Functioning and the Clinical Method", *Psychological Review* 62, no. 4 (1955): s. 255-262; Natalia Karelaia ve Robin M. Hogarth, "Determinants of Linear Judgment: A Meta-Analysis of Lens Model Studies", *Psychological Bulletin* 134, no. 3 (2008): s. 404-426.

134. Paul E. Meehl, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, Minneapolis: University of Minnesota Press, 1954.

136. *Freud'un bir fotoğrafı*: Paul E. Meehl, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*, Northvale, NJ: Aronson, 1996: s. Ön Söz.

136-137. *birçok alanda bilgili bir âlimdi*: "Paul E. Meehl", edit: Ed Lindzey, *A History of Psychology in Autobiography*, 1989.

137. *"Büyük ve sürekli"*: Paul E. Meehl", edit: Ed Lindzey, *A History of Psychology in Autobiography*, Washington, DC: American Psychological Association, 1989: s. 362.

137. *2000 yılında... gözden geçirilmesi*: William M. Grove ve diğerleri, "Clinical Versus Mechanical Prediction: A Meta-Analysis", *Psychological Assesment* 12, no. 1 (2000): s. 19-30.

137. *"özel" bilgilere erişebilmişlerdi*: William M. Grove ve Paul E. Meehl, "Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy", *Psychology, Public Policy, and Law* 2, no. 2 (1996): s. 293-323.

138. *1960'ların sonunda*: Lweis Goldberg, Man Versus Model of Man: A Rationale, plus Some Evidence, for a Method of Improving on Clinical Inferences", *Psychological Bulletin* 73, no. 6 (1970): s. 422-432.

138. *hiç böyle bir şey yapmasalar da*: Milton Friedman ve Leonard J. Savage, "The Utility Analysis of Choices Involving Risk", *Journal of Political Economy* 56, no. 4 (1948): s. 279-304.

138. *bu korelasyon*: Karelaia ve Hogarth, "Determinants of Linear Judgment", s. 411, Tablo 1.

139. *eski bir taklidi*: Nancy Wiggins ve Eileen S. Kohen, "Man Versus Model of Man Revisited: Thee Forecasting of Graduate School Success", *Journal of Personality and Social Psychology* 19, no. 1 (1971): s. 100-106.

139. *elli yıllık araştırmanın*: Karelaia ve Hogarth, "Determinants of Linear Judgment".

141. *kestirimci isabetliliğinizde iyileşme*: Kestiricinin kusurlu güvenilirliği için bir korelasyon katsayısının düzeltilmesi, *daraltma için düzeltme* olarak bilinir. Formül şudur Düzeltilmiş $r_xy = r_xy/\sqrt{r_{xx}}$, r_{xx} güvenilirlik katsayısı olmak üzere (gerçek varyansın kestiricinin gözlemlenen varyansına oranı).

142. *Martin Yu ve Nathan Kuncel'in bir çalışması*: Yu ve Kuncel, "Judgmental Consistency".

142. *rastgele formüller*: Eşit ağırlık ve rastgele ağırlık modellerini bir sonraki bölümde daha detaylı ele alıyoruz. Ağırlıklar bir dizi küçük sayıyla sınırlı ve doğru işareti alacak şekilde sınırlılar.

BÖLÜM 10

145. *kat kat üstün*: Robyn M. Dawes ve Bernard Corrigan, "Linear Models in Decision Making", *Psychological Bulletin* 81, no.2 (1974): s. 95-106. Dawes ve Corrigan ayrıca rastgele ağırlık kullanmayı önerdiler. Bölüm 9'da anlatılan yönetimsel performans öngörüsü bu fikrin uygulamasıdır.

146. *"istatistiki sezgiye ters"*: Jason Dana, "What Makes Improper Linear Models Tick?", edit: Joachim I. Krueger, *Rationality and Social Responsibility: Essays in Honor of Robyn M. Dawes*, New York: Psychology Press, 2008: s. 71-89, s. 73.

147. *benzer sonuçlar*: Jason Dana ve Robyn M. Dawes, "The Superiority of Simple Alternatives to Regression for Social Sciences Prediction", *Journal of Educational and Behavior Statistics* 29 (2004): s. 317-331; Dana, "What Makes Improper Linear Models Tick?"

147. *"Benim için Fark Etmez."*: Howard Wainer, "Estimating Coefficients in Linear Models: It Don't Make No Nevermind", *Psychological Bulletin* 83 no. 2 (1976): s. 213-217.

147. *"... ihtiyacımız yok."*: Dana, "What Makes Improper Linear Models Tick?", s. 72.

148. *Sonuçla korelasyonu*: Martin C. Yu ve Nathan R. Kuncel, "Pushing the Limits for Judgmental Consistency: Comparing Random Weighting Schemes with Expert Judgments", *Personnel Assessment and Decisions* 6, no. 2 (2020): s. 1-10, Bir önceki bölümde olduğu gibi bildirilen korelasyon incelenen üç örneğin ağırlıksız ortalamasıdır. Karşılaştırma örneklerin üçünde de geçerlidir. Klinik uzman yargının geçerliliği 0,17, 0,16 ve 0,13'tü ve eşit ağırlık modellerinin geçerliliği de sırasıyla 0,19, 0,33 ve 0,22 idi.

148. *"sağlıklı bir güzellik"*: Robyn M. Dawes, "Thee Robust Beauty of Improper Linear Models in Decision Making", *American Psychologist* 34, no. 7(1979): s. 571-582.

148. *"Bütün hile"*: Dawes ve Corrigan, "Linear Models in Decision Making", s. 105.

149. *Bir grup araştırmacı*: Jongbin Jung, Conner Concannon, Ravi Shroff, Sharad Goel ve Daniel G. Goldstein, "Simple Rules to Guide Expert Classifications", *Journal of the Royal Statistical Society, Statistics in Society*, no. 183 (2020): s. 771-800.

150. *başka bir araştırma ekibi*: Julia Dressel ve Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism", *Science Advances* 4, no. 1 (2018): s. 1-6.

150. *Sadece iki girdi*: Bu iki örnek son derece küçük değişken setlerine (ve kefalet modeli örneğinde, modeli bir zarfın arka yüzünde yapılan bir hesaplamaya dönüştüren bir yuvarlama yöntemiyle elde edilen doğrusal ağırlıkların tahminine) dayanan doğrusal modellerdir. Bir diğer "belirsiz model" de tek bir kestiriciyi hesaba katan ve diğerlerini yok sayan *tek değişken kuralıdır.* Bakınız Peter M. Todd ve Gerd Gigerenzer, "Précis of Simple Heuristics That Make Us Smart", *Behavioral and Brain Sciences* 23, no. 5 (2000): s. 727-741.

150. *kanıtlarla sabitlenmiş*: P. Gendreau, T. Little ve C. Goggin, "A Meta-Analysis of the Predictors of Adult Offender Recidivism: What Works!", *Criminology* 34 (1996).

150. *çok büyük veri setleri*: Bu bağlamda boyut gözlemlerin sayısının kestiricilere oranı olarak anlaşılmalıdır. Dawes, "Sağlıklı Güzellik", ideal ağırlıkların çapraz doğrulamada birim ağırlıklardan daha iyi sonuç vermesi için 15'e veya 20'ye 1 kadar yüksek olması gerektiğini öne sürdü. Dana ve Dawes, "Basit Alternatiflerin Üstünlüğü", çok daha fazla vaka incelemesi kullanarak çıtayı 1'e 100'e yükseltti.

151. *bir başka ekip*: J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig ve S. Mullainathan, "Human Decisions and Machine Predictions", *Quarterly Journal of Economics* 133 (2018): s. 237-293.

152. *bir yapay öğrenme algoritmasını... eğittiler*: Algoritma eğitim verilerinin bir alt kümesiyle eğitildi ve daha sonra farklı, rastgele seçilmiş bir alt küme üstünde sonuçları öngörme becerisi açısından değerlendirildi.

153. *Yapay öğrenmeyle inşa edilen model*: Kleinberg ve diğerleri, Human Decisions", s. 16.

153-154. *"Sistem gürültüsü... içeriyordu."*: Gregory Stoddard, Jens Ludwig ve Sendhil Mullainathan, yazarlarla e-posta alışverişleri, Haziran-Temmuz 2020.

154. *yazılım mühendislerinin işe alınmasını*: B. Cowgill, "Bias and Productivity in Humans and Algorithms: Theory and Evidence from Résumé Screening", Smith Entrepreneurship Research Conference, College Park, MD sunulan makale 21 Nisan 2018.

157. *1996 yılında birlikte kaleme aldıkları bir makalede*: William M. Grove ve Paul E. Meehl, "Comparative Efficiency of Informal (Subjective, Impressionistic) and Formal (Mechanical, Algorithmic) Prediction Procedures: The Clinical-Statistical Controversy", *Psychology, Public Policy, and Law* 2, no. 2 (1996): s. 293-323.

157. *algoritmayı yeğliyorlar*: Jennifer M. Logg, Julia A. Minson ve Don A. Moore, "Algorithm Appreciation: People Prefer Algorithmic to Human Judgment", *Organizational Behavior and Human Decision Processes* 151 (Nisan 2018): s. 90-103.

157. *hatalar yaptığını görür görmez*: B. J. Dietvorst, J. P. Simmons ve C. Massey, "Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err", *Journal of Experimental Psychology General* 144 (2015): s. 114-126. Bakınız ayrıca A. Prahl ve L. Van Swol. Understanding Algorithm Aversion: When Is Advice from Automation Discounted?", *Journal of Forecasting* 36 (2017): s. 691-702.

157. *Bu beklenti çiğnenirse*: G. Pierce, H. P. Beck ve L. A. Dawe, "The Perceived Utility of Human and Automated Aids in a Visual Detection Task", *Human Factors: The Journal of the Human Factors and Ergonomics Society* 44, no. 1 (2002): s. 79-94; K. A. Hoff ve M. Bashir, "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust", *Human Factors: The Journal of the Human Factors and Ergonomics Society* 57, no. 3 (2015): s. 407-434 ve P. Madhavan ve D. A. Wiegmann, "Similarities and Differences Between Human-Human and Human-Automation Trust: An Integrative Review", *Theoretical Issues in Ergonomics Science* 8, no. 4 (2007): s. 277-301.

BÖLÜM 11

137. *Yönetimsel karar alma konulu araştırmalar*: E. Dane ve M. G. Pratt, "Exploring Intuition and Its Role in Managerial Decision Making", *Academy of Management Review* 32, no. 1 (2007): s. 33-54; Çınla Akıncı ve Eugene Sadler-Smith, "Intuition in Management Research: A Historical Review", *International Journal of Management Reviews* 14 (2012): s. 104-122 ve Gerard P. Hodgkinson ve diğerleri, "Intuition in Organizations: Implications for Strategic Management", *Long Range Planning 42* (2009): s. 277-297.

159. *bir çalışma*: Hodgkinson ve diğerleri, "Intuition in Organizations", s. 279.

161. *yakın zamanda yapılan bir çalışma*: Nathan Kuncel ve diğerleri, "Mechanical Versus Clinical Data Combination in Selection and Admissions Decisions: A Meta-Analysis", *Journal of Applied Psychology* 98, no. 6 (2013): s. 1060-1072. Personel kararları hakkında daha fazla tartışma için bakınız ayrıca Bölüm 24.

162. *Aşırı öz güven*: Don A. Moore, *Perfectly Confident: How to Calibrate Your Decisions Wisely*, New York: HarperCollins, 2020.

163. *"yorum yaparak ya da tavsiyelerde bulunarak"*: Philip E. Tetlock, *Expert Political Judgment: How Good Is It? How Can We Know?*, Princeton, NJ: Princeton University Press, 2005: s. 239 ve 233.

165. *136 çalışmanın gözden geçirilmesi*: William M. Grove ve diğerleri, "Clinical Versus Mechanical Prediction: A Meta-Analysis", *Psychological Assessment* 12, no. 1 (2000): s. 19-30.

166. *kalp krizi*: Sendhil Mullainathan ve Ziad Obermeyer, "Who Is Tested for Heart Attack and Who Should Be: Predicting Patient Risk and Physician Error", 2019 NBER Working Paper 26168, National Bureau of Economic Research.

168. *fazlasıyla belirsiz olarak algıladıkları durumlarda*: Weston Agor, "The Logic of Intuition: How Top Executives Make Important Decisions", *Organizational Dynamics* 14, no. 3 (1986): s. 5-18; Lisa A. Burke ve Monica K. Miller, "Taking the Mystery Out of Intuitive Decision Making", *Academy of Management Perspectives* 13, no. 4 (1999): s. 91-99.

169. *güvenmeye hazırdırlar*: Poornima Madhavan ve Douglas A. Wiegmann, "Effects of Information Source, Pedigree, and Reliability on Operator Interaction with Decision Support Systems", *Human Factors: The Journal of the Human Factors and Ergonomics Society* 49, no. 5 (2007).

BÖLÜM 12

171. *sıra dışı bir makale*: Matthew J. Salganik ve diğerleri, "Measuring the Predictability of Life Outcomes with a Scientific Mass Collaboration", *Proceedings of the National Academy of Sciences* 117, no. 15 (2020): s. 8398-8403.

172. *toplam örneğin*: Kırılgan Aileler çalışmasındaki ailelerin bazıları bu analizden mahremiyet nedeniyle hariç tutulduğu için, toplamda 4.242 aileyi içeriyordu.

173. *0,22'lik bir korelasyona*: İsabetlilliği puanlamak için, yarışmanın organizatörleri Kısım 1'de anlattığımız ölçümü kullandılar: ortalama karesel hata ya da OKH. Karşılaştırmayı kolaylaştırmak için, her modelin OKH'sini "işe yaramaz" bir öngörü stratejisiyle (her duruma uyan, vakaların eğitim setinin ortalamasından farklı olmadığı öngörüsüyle) karşılaştırmalı değerlendirdiler. Kolaylık olması için sonuçlarını korelasyon katsayılarına çevirdik. OKH ve korelasyon r^2 = (Var (Y) – OKH)

/ Var (Y) ile ilişkilendirildi. Var (Y) sonuç değişkeninin varyansı ve (Var (Y) - OKH) öngörülen sonuçların varyansı olmak üzere.

173. *Sosyal psikolojide yüzyıl içinde yapılmış*: F. D. Richard ve diğerleri, "One Hundred Years of Social Psychology Quantitatively Described", *Review of General Psychology* 7, no. 4 (2003): s. 331-363.

173. *708 çalışmanın bir incelemesi*: Gilles E. Gignac ve Eva T. Szodorai, "Effect Size Guidelines for Individual Differences Researchers", *Personality and Individual Differences* 102 (2016): s. 74-78.

174. *"Araştırmacılar... barışmalılar."*: Bir uyarı yerinde olacaktır: Tasarım gereği, bu çalışma çok büyük olan ama belli sonuçları öngörecek şekilde özel hazırlanmamış mevcut bir tanımlayıcı veri setini kullanır. Bu, Tetlock'un çalışmasındaki uygun gördükleri her bilgiyi kullanma özgürlüğüne sahip olan uzmanlarda önemli bir farktır. Örneğin, veri tabanında olmayan ama makul şekilde toplanabilecek tahliye kestiricilerini teşhis etmek mümkün olabilir. Dolayısıyla, çalışma tahliyelerin ve diğer sonuçların nasıl *aslen* öngörülemez olduğunu değil, çok sayıda sosyal bilimci tarafından kullanılan *bu veri setine göre* nasıl öngörülemez olduklarını ispatlar.

175. *nedensel bir zincir*: Jake M. Hofman ve diğerleri, Prediction and Explanation in Social Systems", *Science* 355 (2017): s. 486-488; Duncan J. Watts ve diğerleri, "Explanation, Prediction, and Causality: Three Sides of the Same Coin?", Ekim 2018, s. 1-14; ayrıca bakınız Center for Open Science https://osf.io/bgwjc.

176. *Akıllarımıza daha doğal olarak gelen*: Yakından ilişkili bir başka ayrım da uzantısal ve uzantısal olmayan ya da kasti düşünmeyi karşılaştırır. Amos Tversky ve Daniel Kahneman, "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment", *Psychological Review* 4 (1983): s. 293-315.

178. *geriye dönük*: Daniel Kahneman ve Dale T. Miller, "Norm Theory: Comparing Reality to Its Alternatives", *Psychological Review* 93, no. 2 (1986): s. 136-153.

179. *Geriye bakış üzerine yapılan klasik araştırmalardan*: Baruch Fischhoff, "An Early History of Hindsight Research", *Social Cognition* 25, no. 1 (2007): s. 10-13 (doi:10,1521/soco.2007.25.1.10); Baruch Fischhoff, "Hindsight Is Not Equal to Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty", *Journal of Experimental Psychology: Human Perception and Performance* 1, no. 3 (1975): s. 288.

180. *Sistem 2*: Daniel Kahneman, *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 2011.

BÖLÜM 13

185. *ilk kırk yılı*: Daniel Kahneman, *Thinking, Fast and Slow*, New York: Farrar, Straus and Giroux, 2011.

187-188. *kanıtlar... gösteriyor*: Bir uyarı gerekiyor. Yargı ön yargılarını inceleyen psikologlar, haklı nedenlerle Şekil 10'da görüldüğü gibi her grupta beş katılımcıyla yetinmez: Yargılar gürültülü olduğu için, her deney grubunun sonuçlarının Şekil 11'in düşündürdüğü gibi yakın şekilde öbeklenmesi çok nadirdir. İnsanlar her ön yargıya açıklıklarında farklılık gösterir ve alakalı değişkenleri *tamamen* yol saymazlar. Örneğin, çok sayıda katılımcıyla *kapsam duyarsızlığının* kusurlu olduğunu neredeyse kesin olarak teyit edebilirsiniz: Gambardi'nin görevden ayrılmasına verilen ortalama olasılık üç yıl için iki yıl için olduğundan birazcık daha yüksek. Yine de *kapsam duyarsızlığının* tanımı uygun çünkü fark olması gerekenin küçük bir dilimi.

189. *çok sayıda deneyin*: Daniel Kahneman ve diğerleri (ed), *Judgment Under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press, 1982: Bölüm 6; Daniel Kahneman ve Amos Tversky, "On the Psychology of Prediction", *Psychological Review* 80, no. 4 (1973): s. 237-251.

191. *Amerikan şirketlerindeki CEO değişiminin*: Bakınız örneğin Steven N. Kaplan ve Bernadette A. Minton, "How Has CEO Turnover Changed?", *International Review of Finance* 12, no. 1 (2012): s. 57-87. Bakınız ayrıca Dirk Jenter ve Katharina Lewellen, "Performance-Induced CEO Turnover", Harvard Law School Forum on Corporate Governance, 2 Eylül 2020; https://corpgov.law.harvard.edu/2020/09/02/performance-induced-ceo-turnover.

193. *kilit bir anında*: J. W. Rinzler, *The Making of Star Wars: Return of the Jedi: The Definitive Story*, New York: Del Rey, 2012: s. 64.

193. *senaryosunun geliştirilmesinin*: Cass Sunstein, *The World According to Star Wars*, New York: HarperCollins, 2016.

194. *seçici ve çarpıtılmış*: Burada yargı başladığında bir peşin hükmün var olduğu basit bir vakayı vurguluyoruz. Aslında, böyle bir peşin hükmün yokluğunda bile, kanıtlar biriktikçe, basitlik ve tutarlılığa meyil nedeniyle, belirli bir sonuca yönelik bir ön yargı gelişebilir. Teyide muhtaç

bir sonuç ortaya çıktıkça, onay ön yargısı yeni kanıtların toplanmasını ve yorumlanmasını o sonucun lehine yönlendirebilir.

195. *akıl yürütme şekli*: Bu gözleme *inanç ön yargısı* diyoruz. Bakınız J. St. B. T. Evans, Julie L. Barson ve Paul Pollard, "On the Conflict between Logic and Belief in Syllogistic Reasoning", *Memory & Cognition* 11, no. 3 (1983): s. 295-306.

195. *Tipik bir örnekte*: Dan Ariely, George Loewenstein ve Drazen Prelec, "'Coherent Arbitrariness': Stable Demand Curves Without Stable Preferences", *Quarterly Journal of Economics* 118, no. 1 (2003): s. 73-105.

195. *pazarlıklarda*: Adam D. Galinsky ve T. Mussweiler, "First Offers as Anchors: The Role of Perspective-Taking and Negotiator Focus", *Journal of Personality and Social Psychology* 81, no. 4 (2001): s. 657-669.

196. *aşırı tutarlılık*: Solomon E. Asch, "Forming Impressions of Personality", *Journal of Abnormal and Social Psychology* 41, no. 3 (1946): s. 258-290; bu fenomeni betimlemek için ilk kez bir dizi sıfatı farklı sıralamalarla kullandı.

197. *aydınlatıcı bir çalışmada*: Steven K. Dallas ve diğerleri, "Don't Count Calorie Labeling Out: Calorie Counts on the Left Side of Menu Items Lead to Lower Calorie Food Choices", *Journal of Consumer Psychology* 29, no. 1 (2019): s. 60-69.

BÖLÜM 14

203. *bir yoğunluk ölçeğiyle bir diğeri arasında*: S. S. Stevens, "On the Operation Known as Judgment", *American Scientist* 54, no. 4 (Aralık 1996): s. 385-401. Bizim eşleştirme terimini kullanımımız Stevens'ın Bölüm 15'te geri döneceğimiz orantı ölçekleriyle sınırlı olan kullanımından daha kapsamlı.

204. *sistematik bir yargı hatasını*: Örneğe ilk kez Daniel Kahneman, *Hızlı ve Yavaş Düşünme*'de değinildi.

205. *tamamen aynı sayıları*: Daniel Kahneman ve Amos Tversky, "On the Psychology of Prediction", *Psychological Review* 80 (1973): s. 237-251.

209. *"Sihirli Sayı Yedi"*: G. A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", *Psychological Review* (1956): s. 63-97.

211. *Karşılaştırmalı ölçeklerde*: R. D. Goffin ve J. M. Olson, "Is It All Relative? Comparative Judgments and the Possible Improvement of Self-Ratings

and Ratings of Others", *Perspectives on Psychological Science* 6 (2011): s. 48-60.

BÖLÜM 15

215. *1998'de raporladığımız*: Daniel Kahneman, David Schkade ve Cass Sunstein, "Shared Outrage and Erratic Awards: The Psychology of Punitive Damages", *Journal of Risk and Uncertainty* 16 (1998): s. 49-86; https://link.springer.com/article/10,1023/A:1007710408413 ve Cass Sunstein, Daniel Kahneman ve David Schkade, "Assessing Punitive Damages (with Notes on Cognition and Valuation in Law)", *Yale Law Journal* 107, no. 7 (Mayıs 1998): s. 2071-2153. Araştırmanın masrafları bir seferlik bir ayarlama olarak Exxon tarafından karşılandı ama şirket ne araştırmacılara ücret ödedi ne veri üstünde kontrolü vardı ne de sonuçlar hakkında akademik dergilerde yayımlanmadan önce bilgi aldı.

215. *"Şüpheye yer..."*: A. Keane ve P. McKeown, *The Modern Law of Evidence*, New York: Oxford University Press, 2014.

215. *"olması pek mümkün olmayan"*: Andrew Mauboussin and Michael J. Mauboussin, "If You Say Something Is 'Likely,' How Likely Do People Think It Is?", *Harvard Business Review*, 3 Temmuz 2018.

216. *new BMW: BMW v. Gore*, 517 U.S. 559 (1996); https://supreme.justia.com/cases/federal/us/517/559.

217. *Öfke duygusu*: Ahlaki yargıda duygunun rolü tartışmaları için bakınız J. Haidt, "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment", *Psychological Review* 108, no. 4 (2001): s. 814-834; Joshua Greene, *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*, New York: Penguin Press, 2014.

220. *Şekil 13 sonuçları gösteriyor*: Bu puanlamalardaki büyük gürültü miktarına bakınca, öfke ve cezalandırıcı niyet yargıları arasındaki öfke hipotezine destek sağlayan çok yüksek korelasyon (0,98) kafanızı karıştırabilir. Korelasyonun yargıların *ortalamaları* arasında hesaplandığını hatırlayınca bu kafa karışıklığı ortadan kaybolur. 100 yargının ortalamasında, gürültü (yargıların standart sapması) 10 faktör azalır. Birçok yargı bir araya toplanınca, gürültü bir faktör olmaktan çıkar. Bakınız Bölüm 21.

222. *yoğunluk oranları*: S. S. Stevens, Psychophysics: *Introduction to Its Perceptual, Neural and Social Prospects*, New York: John Wiley & Sons, 1975.

223. *"tutarlı keyfiyet"*: Dan Ariely, George Loewenstein ve Drazen Prelec, "'Coherent Arbitrariness': Stable Demand Curves Without Stable Preferences", *Quarterly Journal of Economics* 118 no. 1 (2003): s. 73-106.

224. *Dolar miktarlarının sıralamalara dönüştürülmesi*: Sıralamaya dönüştürme yargılar arasındaki mesafe korunmadığı için, bir bilgi kaybına neden olur. Sadece üç vaka olduğunu ve bir jüri üyesinin on milyon, iki milyon ve bir milyon dolar ceza önerdiğini düşünün. Jüri üyesi açıkça, ceza niyeti açısından ilk iki vaka arasında, ikinci ve üçüncü arasında olduğundan daha büyük bir fark olduğunu ifade etmek istemektedir. Ancak bu cezalar sıralamaya dönüştürüldüğünde, fark aynı (sadece bir sıra fark) olacaktır. Bu sorun yargıları standart puanlara dönüştürerek çözülebilir.

BÖLÜM 16

230. *(algıda çok incelenmiş bir süreç)*: R. Blake ve N. K. Logothetis, "Visual competition", *Nature Reviews Neuroscience* 3 (2002): s. 13-21; M. A. Gernsbacher ve M. E. Faust, "The Mechanism of Suppression: A Component of General Comprehension Skill", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 17 (Mart 1991): s. 254-262 ve M. C. Stites ve K. D. Federmeier, "Subsequent to Suppression: Downstream Comprehension Consequences of Noun/Verb Ambiguity in Natural Reading", *Journal of Experimental Psychology: Learning, Memory ve Cognition* 41 (Eylül 2015): s. 1497-1515.

230. *olmamız gerekenden daha fazla emin*: D. A. Moore ve D. Schatz, "The three faces of overconfidence", *Social and Personality Psychology Compass* 11, no. 8 (2017), makale e12331.

232. *Kurumsal itibar araştırması*: S. Highhouse, A. Broadfoot, J. E. Yugo ve S. A. Devendorf, "Examining Corporate Reputation Judgments with Generalizability Theory", *Journal of Applied Psychology* 94 (2009): s. 782-789. Scott Highhouse ve Alison Broadfoot'a bize orijinal verilerini açtıkları için, Julian Parris'e de ek analizleri için teşekkür ederiz.

233. *takımlar oluşturma*: P. J. Lamberson ve Scott Page, "Optimal forecasting groups", *Management Science* 58, no. 4 (2012): s. 805-810. Scott Page'e örüntü gürültüsünün bu kaynağına dikkatimizi çektiği için teşekkür ederiz.

235. *Çok önce yapılan... taramasının*: Allport ve Odbert'in (1936) İngilizcede kişilikle ilgili kelimeler çalışmasından bahsedilen Oliver P. John ve Sanjay Srivastava'nın "The Big-Five Trait Taxonomy: History, Measurement, and

Theoretical Perspectives", edit: L. Pervin ve Oliver P. John, *Handbook of Personality: Theory and Research*, 2. Baskı, New York: Guilford, 1999.

235. *yüksek kabul edilir*: Ian W. Eisenberg, Patrick G. Bissett, A. Zeynep Enkavi ve diğerleri, "Uncovering the structure of self-regulation through data-driven ontology discovery", *Nature Communications* 10 (2019): s. 2319.

236. *fiziksel olarak tehdit altında*: Walter Mischel, "Toward an integrative science of the person", *Annual Review of Psychology* 55 (2004): s. 1-22.

BÖLÜM 17

239. *OKH'nin... ayrıldığını*: Ön yargı ve gürültünün çözümlenmesi konusunda genel bir kural olmamakla birlikte, bu şekildeki orantılar, incelediğimiz bazı örneklerin (gerçek ya da kurgusal) kabaca temsilcisidir. Özellikle bu şekilde ön yargı ve gürültü eşittir (Tıpkı GoodSell'in satış öngörülerinde olduğu gibi). Yatay gürültünün karesi sistem gürültüsünün %37'şini açıklamaktadır (tıpkı ceza tazminatları çalışmasında olduğu gibi). Durum gürültüsünün karesi, gösterildiği gibi, örüntü gürültüsünün karesinin yaklaşık %35'idir.

241. *patent bürolarıyla ilgili*: Bakınız girişteki referanslar. Mark A. Lemley ve Bhaven Sampat, "Examiner Characteristics and Patent Office Outcomes", *Review of Economics and Statistics* 94, no. 3 (2012): s. 817-827. Bakınız ayrıca Iain Cockburn, Samuel Kortum ve Scott Stern, "Are All Patent Examiners Equal? The Impact of Examiner Characteristics", çalışma makalesi 8980, Haziran 2002; www.nber.org/papers/w8980 ve Michael D. Frakes ve Melissa F. Wasserman, "Is the Time Allocated to Review Patent Applications Inducing Examiners to Grant Invalid Patents? Evidence from Microlevel Application Data", *Review of Economics and Statistics* 99, no. 3 (Temmuz 2017): s. 550-563.

241. *çocuk koruma hizmetleri*: Joseph J. Doyle Jr., "Child Protection and Child Outcomes: Measuring the Effects of Foster Care", *American Economic Review* 95, no. 5 (Aralık 2007): s. 1583-1610.

242. *Sığınma talebi kararlarında yargıçların*: Andrew I. Schoenholtz, Jaya Ramji-Nogales ve Philip G. Schrag, "Refugee Roulette: Disparities in Asylum Adjudication", *Stanford Law Review* 60, no. 2 (2007).

242. *yaklaşık 2,8 yıl*: Değer, Bölüm 6'da sunulan etkileşim varyansının, toplam varyansın %23'ü olduğu hesaplamalardan tahmin edilmiştir. Hüküm sürelerinin normal olarak dağıldığı varsayımıyla, rastgele seçilmiş iki gözlem arasındaki ortalama mutlak fark 1.128 SD'dir.

242. *Princeton'da... bir araştırmacı grup*: E. Martinez, B. Labbree, S. Uddenberg ve A. Todorov, "Meaningful 'noise': Comparative judgments contain stable idiosyncratic contributions" (yayımlanmamış).

244. *kefalet yargıçları çalışması*: J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig ve S. Mullainathan, "Human Decisions and Machine Predictions", *Quarterly Journal of Economics* 133 (2018): s. 237-293.

244. *simüle edilen yargıçlara*: Model her yargıç için hem 141.833 vakanın sıralamasını hem de ötesinde kefalet hakkının tanındığı eşiği üretti. Yatay gürültü bu eşiklerin değişkenliğini yansıtırken, örüntü gürültüsü vakaların sıralanmasındaki değişkenliği yansıtır.

244. *sabit örüntü gürültüsü*: Gregory Stoddard, Jens Ludwig ve Sendhil Mullainathan, yazarlarla e-posta alışverişleri, Haziran-Temmuz 2020,

247. *Phil Rosenzweig... ikna edici bir şekilde*: Phil Rosenzweig, *Left Brain, Right Stuff: How Leaders Make Winning Decisions*, New York: PublicAffairs, 2014.

BÖLÜM 18

255. *kalabalığın daha becerikli insanlardan oluşması*: Mannes ve diğerleri, "The Wisdom of Select Crowds", *Journal of Personality and Social Psychology* 107, no. 2 (2014): s. 276-299; Jason Dana ve diğerleri, "The Composition of Optimally Wise Crowds", Decision Analysis 12, no. 3 (2015): s. 130-143.

258-259. *Güven kısayolu*: Briony D. Pulford, Andrew M. Colmna, Eike K. Buabang ve Eva M. Krockow, "The Persuasive Power of Knowledge: Testing the Confidence Heuristic", *Journal of Experimental Psychology: General* 147, no. 10 (2018): s. 1431-1444.

259. *ilişkilendirilir*: Nathan R. Kuncel ve Sarah A. Hezlett, "Fact and Fiction in Cognitive Ability Testing for Admissions and Hiring Decisions", *Current Directions in Psychological Science* 19, no. 6 (2010): s. 339-345.

259. *yorumlamalar hâlâ sürmektedir*: Kuncel ve Hezlett, "Fact and Fiction".

260. *Bir incelemenin ifade ettiği gibi*: Frank L. Schmidt ve John Hunter, "General Mental Ability in the World of Work: Occupational Attainment and Job Performance", *Journal of Personality and Social Psychology* 86, no. 1 (2004): s. 162.

260. *sorumluluk ve... azim*: Angela L. Duckworth, David Weir, Eli Tsukayama ve David Kwok, "Who Does Well in Life? Conscientious Adults Excel in Both Objective and Subjective Success", *Frontiers in Psychology* 3 (Eylül

2012). Azim için bakınız Angela L. Duckworth, Christopher Peterson, Michael D. Matthews ve Dennis Kelly, "Grit: Perseverance and Passion for Long-Term Goals", *Journal of Personality and Social Psychology* 92, no. 6 (2007): s. 1087-1101.

260. *akıcı zekâ*: Richard E. Nisbett ve diğerleri, "Intelligence: New Findings and Theoretical Developments", *American Psychologist* 67, no. 2 (2012): s. 130-159.

260. *"çoğundan daha büyük"*: Schmidt ve Hunter, "Occupational Attainment", s. 162.

260. *0,50 diliminde*: Kuncel ve Hezlett, "Fact and Fiction".

260. *sosyal bilim standartlarında*: Bu korelasyonlar gözlemlenen korelasyonları kriter ve aralık kısıtlamasındaki ölçüm hatası için düzelten meta analizlerden türetilmiştir. Araştırmacılar arasında, bu düzeltmeleri not ortalamasının kestirimci değerini abartıp abartmadığı konusunda bir tartışma mevcuttur. Ancak, bu metodolojik tartışmalar başka kestiricilere de uyduğu için, uzmanlar genellikle not ortalamasının (diğer iş örneği testleriyle birlikte bakınız Bölüm 24) işte başarının erişilebilen iyi kestiricisi olduğu konusunda hemfikirler. Bakınız Kuncel ve Hezlett, "Fact and Fiction".

261. *ortalama GZY'nin altında kalan neredeyse kimseyi*: Schmidt ve Hunter, "Occupational Attainment", s. 162.

261. *en üst %1'lik dilimdeki insanların arasında bile*: David Lubinski, "Exceptional Cognitive Ability: The Phenotype", *Behavior Genetics* 39, no. 4 (2009): s. 350-358.

261. *2013 yılında yapılan bir çalışma Fortune 500 şirketlerinin CEO'larına*: Jonathan Wai, "Investigating America's Elite: Cognitive Ability, Education, and Sex Differences", *Intelligence* 41, no. 4 (2013): s. 203-211.

263. *sorulan diğer sorulardan*: Keela S. Thomson ve Daniel M. Oppenheimer, "Investigating an Alternate Form of the Cognitive Reflection Test", *Judgment and Decision Making* 11, no.1 (2016): s. 99-113.

263. *Daha düşük BYT puanları*: Gordon Pennycook ve diğerleri, "Everyday Consequences of Analytic Thinking", *Current Directions in Psychological Science* 24, no. 6 (2015): s. 425-432.

263. *"sahte haberlere" kanıp kanmayacağını*: Gordon Pennycook ve David G. Rand, "Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning than by Motivated Reasoning", *Cognition* 188 (Haziran 2018): s. 39-50.

263. *insanların akıllı telefonlarını ne kadar çok kullandıklarıyla da*: Nathaniel Barr ve diğerleri, "The Brain in Your Pocket: Evidence That Smartphones Are Used to Supplant Thinking", *Computers in Human Behavior* 48 (2015): s. 473-480.

263. *yansıtıcı düşünce süreci kullanma eğilimini*: Niraj Patel, S. Glenn Baker ve Laura D. Scherer, "Evaluating the Cognitive Reflection Test as a Measure of Intuition/ Reflection, Numeracy, and Insight Problem Solving, and the Implications for Understanding Real-World Judgments and Beliefs", *Journal of Experimental Psychology: General* 148, no. 12 (2019): s. 2129-2153.

263. *biliş ihtiyacı ölçeği*: John T. Cacioppo ve Richard E. Petty, "The Need for Cognition", *Journal of Personality and Social Psychology 42*, no. 1 (1982): s. 116-131.

264. *bilişsel ön yargılara daha az açık olma*: Stephen M. Smith ve Irwin P. Levin, "Need for Cognition and Choice Framing Effects", *Journal of Behavioral Decision Making* 9, no. 4 (1996): s. 283-290.

264. *Spoiler uyarısı*: Judith E. Rosenbaum ve Benjamin K. Johnson, "Who's Afraid of Spoilers? Need for Cognition, Need for Affect, and Narrative Selection and Enjoyment", *Psychology of Popular Media Culture* 5, no. 5 (2007): s. 938-956.

264. *Halpern Eleştirel Düşünme Değerlendirmesi*: Heather A. Butler, "Halpern Critical Thinking Assessment Predicts Real-World Outcomes of Critical Thinking", *Applied Cognitive Psychology* 26, no. 5 (2012): s. 721-729.

265. *tahmin becerisini öngörebilecek*: Uriel Haran, Ilana Ritov ve Barbara Mellers, "The Role of Actively Open-Minded Thinking in Information Acquisition, Accuracy, and Calibration", *Judgment and Decision Making* 8, no. 3 (2013): s. 188-201.

265. *aktif olarak açık görüşlü düşünme*: Haran, Ritov ve Mellers, "Role of Actively Open-Minded Thinking".

265. *öğretilebilir bir beceri*: J. Barın, "Why Teach Thinking? An Essay", *Applied Psychology: An International Review* 42 (1993): s. 191-214; J. Baron, *The Teaching of Thinking: Thinking and Deciding*, 2. Baskı, New York: Cambridge University Press, 1994: s. 127-148.

BÖLÜM 19

268. *merkez bulgularını*: Kusursuz bir inceleme için bakınız Jack B. Soll ve diğerleri, "A User's Guide to Debiasing", edit: Gideon Keren ve George

Wu, *The Wiley Blackwell Handbook of Judgment and Decision Making*, cilt 2, New York: John Wiley & Sons, 2015: s. 684.

269. The Green Book: *The Green Book: Central Government Guidance on Appraisal and Evaluation*, Londra: UK Crown, 2018: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf.

269. dürtmeler: Richard H. Thaler ve Cass R. Sunstein, *Nudge: Improving Decisions About Health, Wealth, and Happiness*, New Heaven, CT: Yale University Press, 2008 (Richard H. Thaler ve Cass R. Sunstein, *Dürtme, Sağlık, Zenginlik ve Mutlulukla İlgili Kararları Uygulamak*, Pegasus Yayınları, 2013).

270. güçlendirme: Ralph Hertwig ve Till Grüne-Yanoff, "Nudging and Boosting: Steering or Empowering Good Decisions", *Perspectives on Psychological Science* 12, no. 6 (2017).

270. *eğitim faydalıdır*: Geoffrey T. Fong ve diğerleri, "The Effects of Statistical Training on Thinking About Everyday Problems", *Cognitive Psychology* 18, no. 3 (1986): s. 253-292.

270. *aşırı öz güvenden*: Willem A. Wagenaar ve Gideon B. Keren, "Does the Expert Know? The Reliability of Predictions and Confidence Ratings of Experts," *Intelligent Decision Support in Process Environments* (1986): s. 87-103.

271. *hataların sayısını azalttı*: Carey K. Morewedge ve diğerleri, "Debiasing Decisions: Improved Decision Making with a Single Training Intervention", *Policy Insights from the Behavioral and Brain Sciences* 2, no. 1 (2015): s. 129-140.

271. *öğrendiklerini... uyguladığını*: Anne-Laure Sellier ve diğerleri, "Debiasing Training Transfers to Improve Decision Making in the Field", *Psychological Science* 30 no. 9 (2019): s. 1371-1379.

273. Ön yargı kör noktası: Emily Pronin ve diğerleri, "The Bias Blind Spot: Perceptions of Bias in Self Versus Others", *Personality and Social Psychology Bulletin* 28, no. 3 (2002): s. 369-381.

274. *etkilemiş olması olası ön yargılar*: Daniel Kahneman, Dan Lovallo ve Olivier Sibony, "Before You Make That Big Decision...", *Harvard Business Review* 89, no. 3 (Haziran 2011): s. 50-60.

274. *kontrol listeleri... uzun bir geçmişe sahipler*: Atul Gawande, *Checklist Manifesto: How to Get Things Right*, New York: Metropolitan Books, 2010.

275. *basit bir kontrol listesi*: Office of Information and Regulatory Affairs, "Agency Checklist: Regulatory Impact Analysis", tarihsiz: www.whitehouse.gov/sites/whitehouse.gov/files/omb/inforeg/inforeg/regpol/RIA_Checklist.pdf.

275. *Ek B'yi hazırladık*: Bu kontrol listesi kısmen Daniel Kahneman ve diğerleri, "Before You Make That Big Decision", *Harvard Business Review*.

275. *benimsenmesini kolaylaştırmak*: Bakınız Gawande, *Checklist Manifesto*.

BÖLÜM 20

279. *"bir insan hatası"*: R. Stacey, "A Report on the Erroneous Fingerprint Individualisation in the Madrid Train Bombing Case", *Journal of Forensic Identification* 54 (2004): s. 707-718.

281. *FBI'ın internet sayfası*: Michael Specter, "Do Fingerprints Lie?", *The New Yorker*, 27 Mayıs 2002. Vurgu sonradan eklendi.

282. *Dror'un ifadesiyle*: I. E. Dror ve R. Rosenthal, "Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts", *Journal of Forensic Science* 53 (2008): s. 900-903.

283. *İlk çalışmada*: E. Dror, D. Charlton ve A. E. Péron, "Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications", *Forensic Science International* 156 (2006): s. 74-78.

283. *İkinci çalışmada*: I. E. Dror ve D. Charlton, "Why Experts Make Errors", *Journal of Forensic Identification* 56 (2006): s. 600-616.

284. *"usta parmak izi uzmanlarının"*: I. E. Dror ve S. A. Cole, "The Vision in 'Blind' Justice: Expert Perception, Judgment, and Visual Cognition in Forensic Pattern Recognition", *Psychonomic Bulletin and Review 17* (2010): s. 161-167, 185. Bakınız ayrıca I. E. Dror, "A Hierarchy of Expert Performance (HEP)", *Journal of Applied Research in Memory and Cognition* (2016): s. 1-6.

284. *ayrı bir çalışmada*: E. Dror ve diğerleri, "Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Expert Consistency and the Effect of a 'Target' Comparison", *Forensic Science International* 208 (2011): s. 10-17.

284. *Daha sonra yapılan, bağımsız bir çalışma*: B. T. Ulery, R. A. Hicklin, M. A. Roberts ve J. A. Buscaglia, "Changes in Latent Fingerprint Examiners' Markup Between Analysis and Comparison", *Forensic Science International* 247 (2015): s. 54-61.

284. *DNA analizleri bile*: I. E. Dror ve G. Hampikian, "Subjectivity and Bias in Forensic DNA Mixture Interpretation", *Science and Justice* 51 (2011): s. 204-208.

284. *Uzmanlar genellikle*: M. J. Saks, D. M. Risinger, R. Rosenthal ve W. C. Thompson, "Context Effects in Forensic Science: A Review and Application of the Science of Science to Crime Laboratory Practice in the United States", *Science Justice Journal of Forensic Science Society* 43 (2003): s. 77-90.

285. *Doğrulayıcı uzman... bildiği için*: Başkan'ın Bilim ve Teknoloji Danışma Konseyi (PCAST), *Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, Washington, DC: Executive Office of the President, PCAST, 2016.

285. *daha sonraki soruşturmasının*: Stacey, "Erroneous Fingerprint".

285. *hayli saygın bağımsız uzman bile*: Dror ve Cole, "Vision in Blind Justice".

286. *ön yargı şelaleleri*: I. E. Dror, "Biases in Forensic Experts", *Science* 360 (2018): s. 243.

286. *fikir değiştirebiliyorlar*: Dror ve Charlton, "Why Experts Make Errors".

286. *2012 yılında... yapılan bir çalışma*: B. T. Ulery, R. A. Hicklin, J. A. Buscaglia ve M. A. Roberts, "Repeatability and Reproducibility of Decisions by Latent Fingerprint Examiners", *PLoS One* 7 (2012).

287. *Masumiyet Projesi*: Masumiyet Projesi, "Overturning Wrongful Convictions Involving Misapplied Forensics", *Misapplication of Forensic Science* 2018): s. 1-7; www.innocenceproject.org/causes/misapplication-forensic-science. Ayrıca bakınız S. M. Kassin, I. E. Dror, J. Kukucka ve L. Butt, "The Forensic Confirmation Bias: Problems, Perspectives, and Proposed Solutions", *Journal of Applied Research in Memory and Cognition* 2 (2013): s. 42-52.

287. *derinlemesine inceleyen*: PCAST, *Report to the President*.

287. *geniş ölçekli çalışmadan*: B. T. Ulery, R. A. Hicklin, J. Buscaglia ve M. A. Roberts, "Accuracy and Reliability of Forensic Latent Fingerprint Decisions", *Proceedings of the National Academy of Sciences* 108 (2011): s. 7733-7738.

288. "çok daha yüksek": (PCAST), *Başkan'a sunulan rapor*, s. 95. Vurgu orijinaline aittir.

288. *Daha sonra Florida'da yapılan bir çalışma*: Igor Pacheco, Brian Cerchiai ve Stephanie Stoiloff, "Miami-Dade Research Study for the Reliability of the ACE-V Process: Accuracy & Precision in Latent Fingerprint Exami-

nations", nihai rapor, Miami-Dade Police Department Forensic Services Bureau, 2014; www.ncjrs.gov/pdffiles1/nij/grants/248534.pdf.

288. *"çoğu vaka çalışmasında"*: B. T. Ulery, R. A. Hicklin, M. A. Roberts ve J. A. Buscaglia, "Factors Associated with Latent Fingerprint Exclusion Determinations", *Forensic Science International* 275 (2017): s. 65-75.

288. *daha az hatalı olumlu teşhis kararı*: R. N. Haber ve I. Haber, "Experimental Results of Fingerprint Comparison Validity and Reliability: A Review and Critical Analysis", *Science & Justice* 54 (2014): s. 375-389.

289. *"ön yargıyı yaratmak... daha kolay"*. Dror, "Hierarchy of Expert Performance", s. 3.

289. *"Disneyland'da çalışacak kadar..."*: M. Leadbetter, editöre mektup, *Fingerprint World* 33 (2007): s. 231.

289. *yargılarını değiştirmeden*: L. Butt, "The Forensic Confirmation Bias: Problems, Perspectives and Proposed Solutions—Commentary by a Forensic Examiner", *Journal of Applied Research in Memory and Cognition* 2 (2013): s. 59-60, vurgu sonradan eklendi.

290. *FBI bile*: Stacey, "Erroneous Fingerprint", s. 713, vurgu sonradan eklendi.

290. *dört yüz... yapılan bir ankette*: J. Kukucka, S. M. Kassin, P. A. Zapf ve I. E. Dror, "Cognitive Bias and Blindness: A Global Survey of Forensic Science Examiners", *Journal of Applied Research in Memory and Cognition* 6 (2017).

291. *doğrusal sıralı ifşadır*: I. E. Dror ve diğerleri, editöre mektup: "Context Management Toolbox: A Linear Sequential Unmasking (LSU) Approach for Minimizing Cognitive Bias in Forensic Decision Making", *Journal of Forensic Science* 60 (2015): s. 1111-1112.

BÖLÜM 21

294. *resmî bürolar*: Jeffrey A. Frankel, "Over-optimism in Forecasts by Official Budget Agencies and Its Implications", ön rapor 17239, Ulusal Ekonomik Araştırmalar Bürosu, Aralık 2011; www.nber.org/papers/w17239.

295. *aşırı öz güvenli olma eğilimindedirler*: H. R. Arkes, "Overconfidence in Judgmental Forecasting", edit: Jon Scott Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners*, cilt. 30, International Series in Operations Research & Management Science, Boston: Springer, 2001.

295. *devam etmekte olan üç aylık bir anket*: Itzhak Ben-David, John Graham ve Campell Harvey, "Managerial Miscalibration", *The Quarterly Journal of Economics* 128, no. 4 (Kasım 2013): s. 1547-1584.

260, *"güvensizliğin... bir hata kaynağı"*: T. R. Stewart, "Improving Reliability of Judgmental Forecasts", edit: Jon Scott Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners*, cilt 30, International Series in Operations Research & Management Science, Boston: Springer, 2001: s. 82 (bundan sonra *Principles of Forecasting* olarak geçecek).

295. *Hukuk profesörlerinden Yüksek Mahkeme'nin*: Theodore W. Ruger, Pauline T. Kim, Andrew D. Martin ve Kevin M. Quinn, "The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decision-Making", *Columbia Law Review* 104 (2004): s. 1150-1209.

295. *hava kirliliği düzenlemesinin*: Cass Sunstein, "Maximin", *Yale Journal of Regulation* (müsvedde 3 Mayıs 2020); https://papers.ssrn.com/solpapers.cfm?abstract_id=3476250.

295. *birçok örneğini*: Pek çok örnek için bakınız Armstrong, *Principles of Forecasting.*

296. *Çok sayıda tahmini... ortalamalarını almaktır*: Jon Scott Armstrong, "Combining Forecasts", *Principles of Forecasting*, s. 417-439.

296. çoğu... *daha iyi bir performans*: "Improving Reliability of Judgmental Forecasts", *Principles of Forecasting*, s. 417-439.

297. *%12,5'lik bir ortalamayla*: Armstrong, "Combining Forecasts".

297. *seçme kalabalık*: "The Wisdom of Select Crowds", *Journal of Personality and Social Psychology* 107, no. 2 (2014): s. 276-299.

297. *tahmin piyasalarının çok iyi iş çıkardığı görülmüştür*: Justin Wolfers ve Eric Zitzewitz, "Prediction Markets", *Journal of Economic Perspectives* 18 (2004): s. 107-126.

297. *tahmin piyasalarından faydalanır*: Cass R. Sunstein ve Reid Hastie, *Wiser: Getting Beyond Groupthink to Make Groups Smarter*, Boston: Harvard Business Review Press, 2014.

297. *Delphi yöntemi*: Gene Rowe ve George Wright, "The Delphi Technique as a Forecasting Tool: Issues and Analysis", *International Journal of Forecasting* 15 (1999): s. 353-375. Bakınız ayrıca Dan Bang ve Chris D. Frith, "Making Better Deci sions in Groups", *Royal Society Open Science* 4, no. 8 (2017).

297. *yerleştirilmesi zor olabilir*: R. Hastie, "Review Essay: Experimental Evidence on Group Accuracy", edit: B. Grofman ve G. Guillermo, *Information Pooling and Group Decision Making*, Greenwich, CT: JAI Press, 1986: s. 129-157.

297. mini-Delphi: Andrew H. Van De Ven ve André L. Delbecq, "The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes", *Academy of Management Journal* 17 no. 4 (2017).

301. *"ortalama bir... daha iyi"*: *Superforecasting*, s. 95.

303. *"en güçlü kestirici"*: *Superforecasting*, s. 231.

303. *"Dene, başarısız ol, analiz et"*: *Superforecasting*, s. 273.

305. *sofistike bir istatistik tekniği*: Ville A. Satopää, Marat Salikhov, Philip E. Tetlock ve Barb Mellers, "Bias, Information, Noise: The BIN Model of Forecasting", 19 Şubat 2020, s. 23; https://dx.doi.org/10, 2139/ssrn.3540864.

306. *"Müdahaleler, isabeti hızlı bir şekilde güçlendirdikleri durumlarda"*: Satopää ve diğerleri, "Bias, Information, Noise", s. 23.

307. *"Takım olmak, eğitimin aksine"*: Satopää ve diğerleri, s. 22.

307. *"Süper tahminciler başarılarını... borçlu olabilirler"*: Satopää ve diğerleri, s. 24.

307. *hem bağımsız hem tamamlayıcı*: Clintin P. Davis-Stober, David V. Budescu, Stephen B. Broomell ve Jason Dana. "The composition of optimally wise crowds". *Decision Analysis* 12, no. 3 (2015): s. 130-143.

BÖLÜM 22

311. *Tendon dejenerasyonun miktarını belirlemek*: Laura Horton ve diğerleri, "Development and Assessment of Inter- and Intra-Rater Reliability of a Novel Ultrasound Tool for Scoring Tendon and Sheath Disease: A Pilot Study", *Ultrasound* 24, no. 3 (2016): s. 134; www.ncbi.nlm.nih.gov/pmc/articles/PMC5105362.

311. *Patologlar... kalın iğne biyopsilerini*: Laura C. Collins ve diğerleri, "Diagnostic Agreement in the Evaluation of Image-guided Breast Core Needle Biopsies", *American Journal of Surgical Pathology* 28 (2004): s. 126; https://journals.lww.com/ajsp/Abstract/2004/01000/Diagnostic_Agreement_in_the_Evaluation_of.15.aspx.

311. *Hızlı antijen sonucu olmadığında*: Julie L. Fierro ve diğerleri, "Variability in the Diagnosis and Treatment of Group A Streptococcal Pharyngitis by Primary Care Pediatricians", *Infection Control and Hospital Epidemiology* 35, no. S3 (2014): s. S79; www.jstor.org/stable/10.1086/677820.

311. *diyabetli kabul edilirsiniz*: Diyabet Testleri, Centers for Disease Control and Prevention; https://www.cdc.gov/diabetes/basics/getting-tested.html (15 Ocak 2020).

312. *bazı hastanelerde ikinci*: Joseph D. Kronz ve diğerleri, "Mandatory Second Opinion Surgical Pathology at a Large Referral Hospital", *Cancer* 86 (1999): s. 2426; https://onlinelibrary.wiley.com/doi/full/10,1002(SICI)1097-0142(19991201)86:11%3C2426::AID-CNCR34%3E3,0,CO;2-3.

312. *Dartmouth Atlas Projesi... adadı*: Materyalin büyük kısmı çevrim içi bulunabilir; kitap boyutunda bir özeti: Dartmouth Medical School, *The Quality of Medical Care in the United States: A Report on the Medicare Program; the Dartmouth Atlas of Health Care* 1999, American Hospital Publishers, 1999.

312. *benzer sonuçlar gözleniyor*: Bakınız örneğin OECD: *Geographic Variations in Health Care: What Do We Know and What Can Be Done to Improve Health Sytem Performance?* (Paris: OECD Publishing, 2014), s. 137-169; Michael P. Hurley ve diğerleri, "Geographic Variation in Surgical Outcomes and Cost Between the United States and Japan", *American Journal of Managed Care* 22 (2016): s. 600; www.ajmc.com/journals/issue/2016/2016-cilt22-n9/geographic-variation-in-surgical- outcomes-and-cost-between-the-united-states-and-japan ve John Appleby, Veena Raleigh, Francesca Frosini, Gwyn Bevan, Haiyan Gao ve Tom Lyscom, *Variations in Health Care: The Good, the Bad and the Inexplicable*, Londra: The King's Fund, 2011; www.kingsfund.org.uk/sites/default/files/Variations-in -health-care-good-bad-inexplicable-report-The-Kings-Fund -April-2011.pdf.

313. *zatürree teşhisleri üstünde yapılan bir incelemede*: David C. Chan Jr. ve diğerleri, "Selection with Variation in Diagnostic Skill: Evidence from Radiologists", Ulusal Ekonomik Araştırmalar Bürosu, NBER Ön Rapor No. 26467, Kasım 2019; www.nber.org/papers/w26467.

313. *Başka alanlarda olduğu gibi burada da, eğitim*: P. J. Robinson, "Radiology's Achilles' Heel: Error and Variation in the Interpretation of the Röntgen Image", *British Journal of Radiology* 70 (1997): s. 1085; www.ncbi.nlm.nih.gov/pubmed/9536897. İlgili bir çalışma olarak Yusuke Tsugawa ve diğerleri, "Physician Age and Outcomes in Elderly Patients in Hospital in the US: Observational Study", *BMJ 357* (2017); www.bmj.com/content/357/bmj.j1797 doktorların gelirinin eğitimden uzaklaştıkları zaman gittikçe kötüleştiğini ortaya koyuyor. Meslekte geçirilen yıllardan gelen

tecrübe ile en son kanıtlara ve kılavuzlara aşinalık arasında bir alışveriş söz konusu. Araştırma en iyi sonuçların uzmanlıktan sonraki ilk yıllarındaki doktorlardan, o kanıtlar zihinlerindeyken çıktığını gösteriyor.

313. *Örneğin, radyologlar*: Robinson, "*Radiology's Achilles' Heel*".

313. *kappa istatistiğiyle*: Pratikte nadir olsa da korelasyoyon katsayısı gibi, kappa negatif olabilir. Farklı kappa istatistiklerinin anlamının bir özelliği: "hafif (K= 0,00-0,20), makul (K= 0,21-0,40), orta (K = 0,41-0,60) ve önemli (K= 0,61- 0,80) ve neredeyse kusursuz (K > 0,80)" (Ron Wald, Chaim M. Bell, Rosane Nisenbaum, Samuel Perrone, Orfeas Liangos, Andreas Laupacis ve Bertrand L. Jaber, "Interobserver Reliability of Urine Sediment Interpretation", *Clinical Journal of the American Society of Nephrology* 4, no. 3 [Mart 2009]: s. 567-571; https://cjasn.asnjournals.org/content/4/3/567).

314. *ilaç etkileşimlerinin*: Howard R. Strasberg ve diğerleri, "Inter-Rater Agreement Among Physicians on the Clinical Significance of Drug-Drug Interactions", AMIA Yıllık Sempozyum Bildirileri (2013): s. 1325; www.ncbi.nlm.nih.gov /pmc/articles/PMC3900147.

314. *Fakat nefrologlar*: Wald ve diğerleri, "Interobserver Reliability of Urine Sediment Interpretation"; https://cjasn.asnjournals.org/content/4/3/567.

314. *Bir meme lezyonunun*: Juan P. Palazzo ve diğerleri, "Hyperplastic Ductal and Lobular Lesions and Carcinomas in Situ of the Breast: Reproducibility of Current Diagnostic Criteria Among Community- and Academic-Based Pathologists", *Breast Journal* 4 (2003): s. 230; www.ncbi.nlm.nih.gov/pubmed/21223441.

314. *Memede proliferatif lezyon*: Rohit K. Jain ve diğerleri, "Atypical Ductal Hyperplasia: Interobserver and Intraobserver Variability", *Modern Pathology* 24 (2011): s. 917; www.nature.com/articles/modpathol201166.

314. *Spinal stenozis derecesini*: Alex C. Speciale ve diğerleri, "Observer Variability in Assessing Lumbar Spinal Stenosis Severity on Magnetic Resonance Imaging and Its Relation to Cross-Sectional Spinal Canal Area", *Spine* 27 (2002): s. 1082; www.ncbi.nlm.nih.gov/pubmed/12004176.

314. *Kalp hastalıkları... başlıca*: Hastalık Kontrol ve Önleme Merkezleri, "Heart Disease Facts"; www.cdc.gov/heartdisease/facts.htm, 16 Haziran 2020.

315. *Eski bir çalışma vakaların %31'inde*: Timothy A. DeRouen ve diğerleri, "Variability in the Analysis of Coronary Arteriograms", *Circulation* 55 (1977): s. 324; www.ncbi.nlm.nih.gov/pubmed/832349.

315. *çarpıcı görüş farklılıkları*: Olaf Buchweltz ve diğerleri, "Interobserver Variability in the Diagnosis of Minimal and Mild Endometriosis", *European Journal of Obstetrics & Gynecology and Reproductive Biology* 122 (2005): s. 213; www.ejog.org/article/S0301-2115(05)00059-X/pdf.

315. *TB teşhisindeki değişkenlik*: Jean-Pierre Zellweger ve diğerleri, "Intraobserver and Overall Agreement in the Radiological Assessment of Tuberculosis", *International Journal of Tuberculosis & Lung Disease* 10 (2006): s. 1123; www.ncbi.nlm.nih.gov/pubmed/17044205. "Makul" değerlendiriciler arası anlaşma için bakınız Yanina Balabanova ve diğerleri, "Variability in Interpretation of Chest Radiographs Among Russian Clinicians and Implications for Screening Programmes: Observational Study", *BMJ* 331 (2005): s. 379; www.bmj.com/content/331/7513/379.short.

315. *farklı ülkelerin radyologları*: Shinsaku Sakurada ve diğerleri, "Inter-Rater Agreement in the Assessment of Abnormal Chest X-Ray Findings for Tuberculosis Between Two Asian Countries", *BMC Infectious Diseases* 12, Madde 31 (2012); https://bmcinfectdis.biomedcentral.com/articles/10, 1186/1471-2334-12-31.

316. *gözden geçiren sekiz patolog*: Evan R. Farmer ve diğerleri, "Discordance in the Histopathologic Diagnosis of Melanoma and Melanocytic Nevi Between Expert Pathologists", *Human Pathology* 27 (1996): s. 528; www.ncbi.nlm.nih.gov/pubmed/=8666360.

316. *Bir onkoloji merkezinde yapılan bir başka çalışmada*: Alfred W. Kopf, M. Mintzis ve R. S. Bart, "Diagnostic Accuracy in Malignant Melanoma", *Archives of Dermatology* 111 (1975): s. 1291; www.ncbi.nlm.nih.gov/pubmed/1190800.

316. *Çalışmanın yazarları... sonucuna vardılar*: Maria Miller ve A. Bernard Ackerman, "How Accurate Are Dermatologists in the Diagnosis of Melanoma? Degree of Accuracy and Implications", *Archives of Dermatology* 128 (1992): s. 559; https://jamanetwork.com/journals/jamadermatology/fullarticle/554024.

316. *Benzer bir şekilde hatalı olumlu*: Craig A. Beam ve diğerleri, "Variability in the Interpretation of Screening Mammograms by US Radiologists", *Archives of Internal Medicine* 156 (1996): s. 209; www.ncbi.nlm.nih.gov/pubmed/8546556.

316. *Radyologlar bazen*: P. J. Robinson ve diğerleri, "Variation Between Experienced Observers in the Interpretation of Accident and Emergency

Radiographs", *British Journal of Radiology* 72 (1999): s. 323; www.birpublications.org/doi/pdf/10,1259/bjr.72.856.10474490.

316. *Anjiyogramlardaki tıkanıklık derecesini*: Katherine M. Detre ve diğerleri, "Observer Agreement in Evaluating Coronary Angiograms", *Circulation* 52 (1975): s. 979; www.ncbi.nlm.nih.gov/pubmed/1102142.

316-317. *Belirsiz kriterleri... içeren alanlarda*: Horton ve diğerleri, "Inter- and Intra-Rater Reliability"; Megan Banky ve diğerleri, "Inter- and Intra-Rater Variability of Testing Velocity When Assessing Lower Limb Spasticity", *Journal of Rehabilitation Medicine* 51 (2019); www.medicaljournals.se/jrm/content/abstract/10,2340/165019077-2496.

317. *içermeyen bir başka çalışma*: Esther Y. Hsiang ve diğerleri, "Association of Primary Care Clinic Appointment Time with Clinician Ordering and Patient Completion of Breast and Colorectal Cancer Screening", *JAMA Network Open* 51 (2019); https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2733171.

317. *Yorgunluğun... bir diğer rolü*: Hengchen Dai ve diğerleri, "The Impact of Time at Work and Time Off from Work on Rule Compliance: The Case of Hand Hygiene in Health Care", *Journal of Applied Psychology* 100 (2015): s. 846; www.ncbi.nlm.nih.gov/pubmed/25365728.

317. *çok büyük bir katkı*: Ali S. Raja, "The HEART Score Has Substantial Interrater Reliability", *NEJM Journal Watch*, 5 Aralık 2018; www.jwatch.org/na47998/2018/12/05/heart-score-has-substantial-interrater-reliability (Colin A. Gershon ve diğerleri, "Inter-rater Reliability of the HEART Score", *Academic Emergency Medicine* 26 [2019]: s. 552).

318. *bahsettiğimiz gibi, eğitim*: Jean-Pierre Zellweger ve diğerleri, "Intra-observer and Overall Agreement in the Radiological Assessment of Tuberculosis", *International Journal of Tuberculosis & Lung Disease* 10 (2006): s. 1123; www.ncbi.nlm.nih.gov/pubmed/17044205; İbrahim Abubakar ve diğerleri "Diagnostic Accuracy of Digital Chest Radiography for Pulmonary Tuberculosis in a UK Urban Population", *European Respiratory Journal* 35 (2010): s. 689; https://erj.ersjournals.com/content/35/3/689.short.

318. *yargılarının bir araya toplanması da öyle*: Michael L. Barnett ve diğerleri, "Comparative Accuracy of Diagnosis by Collective Intelligence of Multiple Physicians vs Individual Physicians", *JAMA Network Open* 2 (2019): s. e19009; https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2726709; Kimberly H. Allison ve diğerleri, "Understanding Diagnostic Variability in Breast Pathology: Lessons Learned from an

Expert Consensus Review Panel", *Histopathology* 65 (2014): s. 240; https://onlinelibrary.wiley.com/doi/abs/10,1111/his.12387.

318. *Bunların en iyilerinin... görüldü*: Babak Ehteshami Bejnordi ve diğerleri, "Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women with Breast Cancer", *JAMA 318* (2017): s. 2199; https://jamanetwork.com/journals/jama/fullarticle/2665774.

318. *derin öğrenme algoritmaları*: Varun Gulshan ve diğerleri, "Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs", *JAMA* 316 (2016): s. 2402; https://jamanetwork.com/journals/jama/fullarticle/2588763.

318. *YZ artık... en az*: Mary Beth Massat, "A Promising Future for AI in Breast Cancer Screening", *Applied Radiology* 47 (2018): s. 22; www.appliedradiology.com/articles/a-promising-future-for-ai-in-breast-cancer-screening; Alejandro Rodriguez-Ruiz ve diğleri, "Stand-Alone Artificial Intelligence for Breast Cancer Detection in Mammography: Comparison with 101Radiologists", *Journal of the National Cancer Institute* 111 (2019): s. 916; https://academic.oup.com/jnci/advance-article-abstract/doi/10,1093/jnci/djy222/5307077.

319. Tablo 3, Apgar Score, Medline Plus; https:///medlineplus.gov/ency/article/003402.htm (4 Şubat 2020).

320. *Apgar puanlamasının çok az gürültü ürettiğini*: "The Interrater Reliability of Apgar Scores at 1 and 5 Minutes", *Journal of Investigative Medicine* 54, no. 1 (2006): s. 293; https://jim.bmj.com/content/54/1/S308.4.

321. *değerlendirme ve puanlama*: Warren J. McIsaac ve diğerleri, "Empirical Validation of Guidelines for the Management of Pharyngitis in Children and Adults", *JAMA* 291 (2004): s. 1587; www.ncbi.nlm.nih.gov/pubmed/15069046.

321. *Bir çalışma, BI-RADS'ın*: Emilie A. Ooms ve diğerleri, "Mammography: Interobserver Variability in Breast Density Assessment", *Breast* 16 (2007): s. 568; www.sciencedirect.com/science/article/abs/pii/S0960977607000793.

321. *Patolojide de*: Frances P. O'Malley ve diğerleri, "Interobserver Reproducibility in the Diagnosis of Flat E Epithelial Atypia of the Breast", *Modern Pathology* 19 (2006): s. 172; www.nature.com/articles/3800514.

321. *Bu nedenle, gürültüyü azaltmak*: bakınız Ahmed Aboraya ve diğerleri, "The Reliability of Psychiatric Diagnosis Revisited", *Psychiatry (Edgmont)* 3 (2006): s. 41; www.ncbi.nlm.nih.gov/pmc/articles/PMC2990547. Bir

inceleme için bakınız N. Kreitman, "The Reliability of Psychiatric Diagnosis", *Journal of Mental Science* 107 (1961): s. 876-886; www.cambridge.org/core/journals/journal-of-mental-science/article/reliability-of-psychiatric-diagnosis/92832FFA170F4FF41189428C6A3E6394.

322. *1964 yılında 91 hasta*: Aboraya ve diğerleri, "Reliability of Psychiatric Diagnosis Revisited", s. 43.

322. *bir ön çaba olarak*: C. H. Ward ve diğerleri, "The Psychiatric Nomenclature: Reasons for Diagnostic Disagreement", *Archives of General Psychiatry* 7 (1962): s. 198.

322. *Gelişimsel eğitimli bir hekim*: Aboraya ve diğerleri, "Reliability of Psychiatric Diagnosis Revisited".

323. *DSM-III... yol açtı*: Samuel M. Lieblich, David J. Castle, Christos Pantelis, Malcolm Hopwood, Allan Hunter Young ve Ian P. Everall, "High Heterogeneity and Low Reliability in the Diagnosis of Major Depression Will Impair the Development of New Drugs", *British Journal of Psychiatry Open* 1 (2015): s. e5-e7; www.ncbi.nlm.nih.gov/pmc/articles/PMC5000492/pdf/bjporcpsych_1_2_e5.pdf.

323. *Ancak el kitabı*: Lieblich ve diğerleri, "High Heterogeneity".

323. *2000 yılındaki hatırı sayılır revizyonundan*: bakınız Elie Cheniaux ve diğerleri, "The Diagnoses of Schizophrenia, Schizoaffective Disorder, Bipolar Disorder and Unipolar Depression: Interrater Reliability and Congruence Between DSM-IV and ICD-10", *Psychopathology* 42 (2009): s. 296-298, özellikle 293 ve Michael Chmielewski ve diğerleri, "Method Matters: Understanding Diagnostic Reliability in DSM-IV and DSM-5", *Journal of Abnormal Psychology* 124 (2015): s. 764,768-769.

323. *"teşhislerin güvenilirliğini artırdığı"*: Aboraya ve diğerleri, "Reliability of Psychiatric Diagnosis Revisited", s. 47.

323. *riski ciddiyetini*: Aboraya ve diğerleri, s. 47.

323. *El kitabının bir diğer versiyonu*: bakınız Chmielewski ve diğerleri, "Method Matters".

323. *Amerikan Psikiyatri Derneği*: bakınız Helena Chmura Kraemer ve diğerleri, "DSM-5: How Reliable Is Reliable Enough?", *American Journal of Psychiatry* 169 (2012): s. 13-15.

323. *psikiyatrlar hâlâ önemli... sergiliyorlar*: Lieblich ve diğerleri, "High Heterogeneity".

323. *psikiyatrların... sıkıntı çektiklerini*: Lieblich ve diğerleri, "High Heterogeneity", s. e-5.

323. *DSM-5 için saha denemeleri*: Lieblich ve diğerleri, s. e-5.
323. *Birtakım saha denemelerine göre: DSM-5*: Lieblich ve diğerleri, s. e-6.
323. *sınırlı başarısının başlıca nedeni*: Aboraya ve diğerleri, "Reliability of Psychiatric Diagnosis Revisited", s. 47.
324. *Bu teklifler (1)*: Aboraya ve diğerleri.
324. *Bir gözlemcinin sözcükleriyle*: Aboraya ve diğerleri.
324. *Tıp mesleğinin daha fazla*: Birtakım değerli uyarıcı notlar için bakınız Christopher Worsham ve Anupam B. Jena, "The Art of Evidence-Based Medicine", *Harvard Business Review*, 30 Ocak 2019; https://hbr.org/2019/01/the-art-of-evidence-based-medicine.

BÖLÜM 23

327. *Bir gazete manşetinin ifade ettiği gibi*: Jena McGregor, "Study Finds That Basically Every Single Person Hates Performance Reviews", *Washington Post*, 27 Ocak 2014.
328. *her yerde rastlanması bu yüzdendir*: Pek çok örgütün yaşadığı dijital dönüşüm burada yeni olasılıklar yaratabilir. Teoride şirketler artık her çalışan hakkında, çok büyük miktarlarda değerli ve gerçek zamanlı bilgi toplayabilirler. Bu veri bazı pozisyonlar için algoritmaya dayalı performans değerlendirmesini mümkün kılar. Ancak biz burada yargının performans ölçümünden tamamen bertaraf edilemeyeceği pozisyonlara odaklanıyoruz. Bakınız E. D. Pulakos, R. Mueller-Hanson ve S. Arad, "The Evolution of Performance Management: Searching for Value", *Annual Review of Organizational Psychology and Organizational Behavior* 6 (2018): s. 249-271.
328. *Çoğu araştırmacı... sonucuna vardı*: S. E. Scullen, M. K. Mount ve M. Goff, "Understanding the Latent Structure of Job Performance Ratings", *Journal of Applied Psychology* 85 (2000): s. 956-970.
328. *Geri kalanı... farklılığın %70 ila %80'i*: Küçük bir öge –bazı çalışmalarda toplam farklılığın %10'u– araştırmacıların *değerlendirici perspektifi* ya da organizasyondaki seviye anlamında *yatay etki* olarak adlandırdığı (bizim *yatay gürültü* olarak tanımladığımız değil) şeydir. Değerlendirici perspektifi (bakış açısı) aynı kişiyi puanlarken, üstlerin sistematik olarak şirketteki mevkidaşlarından, mevkidaşların da astlarından farklılık sergilemesidir. 360 derece puanlama sistemlerinden alınan sonuçların merhametli yorumlanmasında, bunun gürültü olmadığı savunulabilir. Örgütün farklı seviyelerindeki insanlar aynı kişinin performansının farklı

yönlerini görüyorlarsa, o kişi hakkında yargıları sistematik olarak farklılık sergilemeli ve puanlamaları bunu yansıtmalıdır.

328. *sayısız çalışma*: Scullen, Mount ve Goff, "Latent Structure"; C. Viswesvaran, D. S. Ones ve F. L. Schmidt, "Comparative Analysis of the Reliability of Job Performance Ratings", *Journal of Applied Psychology* 81 (1996): s. 557-574. G. J. Greguras ve C. Robie, "A New Look at Within-Source Interrater Reliability of 360-Degree Feedback Ratings", *Journal of Applied Psychology* 81 (1996): s. 557-574; G. J. Greguras, C. Robie, D. J. Schleicher ve M. A. Goff, "A Field Study of the Effects of Rating Purpose on the Quality of Multisource Ratings", *Personnel Psychology* 56 (2003): s. 1-21; C. Viswesvaran, F. L. Schmidt ve D. S. Ones, "Is There a General Factor in Ratings of Job Performance? A Meta-Analytic Framework for Disentangling Substantive and Error Influences", *Journal of Applied Psychology* 90 (2005): s. 108-131 ve B. Hoffman, C. E. Lance, B. Bynum ve W. A. Gentry, "Rater Source Effects Are Alive and Well After All", *Personnel Psychology* 63 (2010): s. 119-151.

329. *"İş performansı... arasındaki ilişkinin"*: K. R. Murphy, "Explaining the Weak Relationship Between Job Performance and Ratings of Job Performance", *Industrial and Organizational Psychology* 1 (2008): s. 148-160, özellikle s. 151.

329. *çalışanın gerçek performansıyla*: Gürültünün kaynaklarıyla ilgili çalışmada, belli çalışanların ya da çalışan kategorilerinin puanlanmasındaki sistematik ön yargılarından doğan vaka gürültüsü olasılığını yok saydık. Performans puanlamalarının değişkenliği üzerine bulabildiğimiz çalışmaların hiçbiri onları dışarıdan değerlendirilmiş "gerçek" bir performansla karşılaştırmamıştı.

329. *insanları "stratejik" olarak puanlayabilir*: E. D. Pulakos ve R. S. O'Leary, "Why Is Performance Management Broken?", *Industrial and Organizational Psychology* 4 (2011): s. 146-164; M. M. Harris, "Rater Motivation in the Performance Appraisal Context: A Theoretical Framework", *Journal of Management* 20 (1994): s. 737-756 ve K. R. Murphy ve J. N. Cleveland, *Understanding Performance Appraisal: Social, Organizational, and Goal-Based Perspectives*, Thousand Oaks, CA: Sage, 1995.

330. *sadece gelişim amaçlı*: Greguras ve diğerleri, "Field Study".

331. *nesnel olarak... öngörmesi*: P. W. Atkins ve R. E. Wood, "Self- Versus Others' Ratings as Predictors of Assessment Center Ratings: Validation Evidence for 360-Degree Feedback Programs", *Personnel Psychology* (2002).

331. *Gereğinden fazla çalışılmış ve üstünde oynanmış anketler*: Atkins ve Wood, "Self- Versus Others' Ratings".

332. *%98*: Olson ve Davis'ten alıntı: Peter G. Dominick, "Forced Ranking: Pros, Cons and Practices", edit: James W. Smither ve Manuel London, *Performance Management: Putting Research into Action*, San Francisco: Jossey-Bass, 2009: s. 411-443.

332. *zorunlu sıralamadır*: Dominick, "Forced Ranking".

332. *performans değerlendirmelerinde de geçerli olduğu*: Barry R. Nathan ve Ralph A. Alexander, "A Comparison of Criteria for Test Validation: A Meta-Analytic Investigation", *Personnel Psychology* 41, no. 3 (1988): s. 517-535.

333. *Şekil 17*: Aktaran Richard D. Goffin ve James M. Olson, "Is It All Relative? Comparative Judgments and the Possible Improvement of Self-Ratings and Ratings of Others", *Perspectives on Psychological Science* 6, no.1 (2011): s. 48-60.

336. *Deloitte*: M. Buckingham ve A. Goodall, "Reinventing Performance Management", *Harvard Business Review*, 1 Nisan 2015: s. 1-16; doi:ISSN:0017-8012.

336. *Bir çalışma*: Corporate Leadership Council, S. Adler ve diğerleri, "Getting Rid of Performance Ratings: Genius or Folly? A Debate", *Industrial and Organizational Psychology* 9 (2016): s. 219-252.

337. *"onca şeye rağmen"*: Pulakos, Mueller-Hanson ve Arad, "Evolution of Performance Management", s. 250.

337. *"performans yönetimi devrimi"*: A. Tavis ve P. Cappelli, "The Performance Management Revolution", *Harvard Business Review*, Ekim 2016: s. 1-17.

338. *kanıtlar... düşündürmektedir*: Frank J. Landy ve James L. Farr, "Performance Rating", *Psychological Bulletin* 87, no. 1 (1980): s. 72-107.

338. *performans puanlama egzersizleri*: D. J. Woehr ve A. I. Huffcutt, "Rater Training for Performance Appraisal: A Quantitative Review", *Journal of Occupational and Organizational Psychology* 67 (1994): s. 189-205; S. G. Roch, D. J. Woehr, V. Mishra ve U. Kieszczynska, "Rater Training Revisited: An Updated Meta-Analytic Review of Frame-of-Reference Training", *Journal of Occupational and Organizational Psychology* 85 (2012): s. 370-395 ve M. H. Tsai, S. Wee ve B. Koh, "Restructured Frame-of-Reference Training Improves Rating Accuracy", *Journal of Organizational Behavior* (2019): s. 1-18; doi:10,1002/job.2368.

338. *Şekil 18*: Sol panel Richard Goffin ve James M. Olson, "Is It All Relative? Comparative Judgments and the Possible Improvement of Self-Ratings

and Ratings of Others", *Perspectives on Psychological Science* 6, no. 1 (2011): s. 48-60.

339. *çalışmaların çoğunluğunun*: Roch ve diğerleri, "Rater Training Revisited".

339. *"yıldız yetenek"*: Ernest Boyle ve Herman Aguinis, "The Best and the Rest: Revisiting the Norm of Normality of Individual Performance", *Personnel Psychology* 65, no. 1 (2012): s. 79-119 ve Herman Aguinis ve Ernest O'Boyle, "Star Performers in Twenty-First Century Organizations", *Personnel Psychology* 67, no. 2 (2014): s. 313-350.

BÖLÜM 24

341. *"nadir... bir durumdur"*: A. I. Huffcutt ve S. S. Culbertson, "Interviews", edit: S. Zedeck, *APA Handbook of Industrial and Organizational Psychology,* Washington, DC: American Psychological Association, 2010: s. 185-203,

342. *bir dereceye kadar sezgisel yargılarına güvenirler*: N. R. Kuncel, D. M. Klieger ve D. S. Ones, "In Hiring, Algorithms Beat Instinct", *Harvard Business Review* 92, no. 5 (2014): s. 32.

342. *"en büyük sorun..."*: R. E. Ployhart, N. Schmitt ve N. T. Tippins, "Solving the Supreme Problem: 100 Years of Selection and Recruitment at the *Journal of Applied Psychology*", *Journal of Applied Psychology* 102 (2017): s. 291-304.

343. *Başka çalışmalar*: M. McDaniel, D. Whetzel, F. L. Schmidt ve S. Maurer, "Meta Analysis of the Validity of Employment Interviews", *Journal of Applied Psychology* 79 (1994): s. 599-616; A. Huffcutt ve W. Arthur, "Hunter and Hunter (1984 Revisited: Interview Validity for Entry-Level Jobs", *Journal of Applied Psychology* 79 (1994): s. 2; F. L. Schmid ve J. E. Hunter, "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings", *Psychology Bulletin* 124 (1998): s. 262-274 ve F. L. Schmidt ve R. D. Zimmerman, "A Counterintuitive Hypothesis About Employment Interview Validity and Some Supporting Evidence", *Journal of Applied Psychology* 89 (2004): s. 553-561. Bu değerlendirmelerin çalışmaların belli alt kümeleri hesaba katıldığında daha yüksek olduğuna dikkat çekmek isteriz. Hele araştırma, var olan idari puanlamalar yerine, özel olarak bu amaç için yaratılmış performans puanlarını kullanıyorsa.

344. *objektif bilgisizlik*: S. Highhouse, "Stubborn Reliance on Intuition and Subjectivity in Employee Selection", *Industrial and Organizati-*

onal Psychology 1 (2008): s. 333-342. D. A. Moore, "How to Improve the Accuracy and Reduce the Cost of Personnel Selection", *California Management Review* 77 (2012): s. 8-17.

344. *kültürel açıdan kendilerine benzer*: L. A. Rivera, "Hiring as Cultural Matching: The Case of Elite Professional Service Firms", *American Sociology Review* 77 (2012): s. 999-1022.

344. *korelasyon ölçümleri*: Schmidt ve Zimmerman, "Counterintuitive Hypothesis"; Timothy A. Judge, Chad A. Higgins ve Daniel M. Cable, "The Employment Interview: A Review of Recent Research and Recommendations for Future Research", *Human Resource Management Review* 10 (2000): s. 383-406 ve A. I. Huffcutt, S. S. Culbertson ve W. S. Weyhrauch, "Employment Interview Reliability: New Meta-Analytic Estimates by Structure and Format", *International Journal of Selection and Assessment* 21 (2013): s. 264-276.

345. *önemli, hem de çok*: M. R. Barrick ve diğerleri, "Candidate Characteristics Driving Initial Impressions During Rapport Building: Implications for Employment Interview Validity", *Journal of Occupational and Organizational Psychology* 85 (2012): s. 330-352; M. R. Barrick, B. W. Swider ve G. L. Stewart, "Initial Evaluations in the Interview: Relationships with Subsequent Interviewer Evaluations and Employment Offers", *Journal of Applied Psychology*: 95 (2010): s. 1163.

345. *El sıkışma şekli*: G. L. Stewart, S. L. Dustin, M. R. Barrick ve T. C. Darnold, "Exploring the Handshake in Employment Interviews", *Journal of Applied Psychology* 93 (2008): s. 1139-1146.

346. *Olumlu ilk izlenim*: T. W. Dougherty, D. B. Turban ve J. C. Callender, "Confirming First Impressions in the Employment Interview: A Field Study of Interviewer Behavior", *Journal of Applied Psychology* 79 (1994): s. 659-665.

346. *Çarpıcı bir deneyde*: J. Dana, R. Dawes ve N. Peterson, "Belief in the Unstructured Interview: The Persistence of an Illusion", *Judgment and Decision Making* 8 (2013): s. 512-520.

348. *İK profesyonellerinin... yeğlediklerini*: Nathan R. Kuncel ve diğerleri, "Mechanical versus Clinical Data Combination in Selection and Admissions Decisions: A Meta-Analysis", *Journal of Applied Psychology* 98, no. 6 (2013): s. 1060-1072.

349. *"sıfır ilişki..."*: Laszlo Bock, Adam Bryant ile röportaj, *The New York Times*, 19 Haziran 2013; bakınız ayrıca Laszlo Bock, *Work Rules!: In-*

sights from Inside Google !at Will Transform How You Live and Lead, New York: Hachette, 2015.

350. *Saygın ve önemli bir yetenek avcısı*: C. Fernández-Aráoz, "Hiring Without Firing", *Harvard Business Review*, 1 Temmuz 1999.

351. *yapılandırılmış davranışsal mülakatlar*: Yapılandırılmış mülakatlara bir rehber olarak bakınız Michael A. Campion, David K. Palmer ve James E. Campion, "Structuring Employment Interviews to Improve Reliability, Validity and Users' Reactions", *Current Directions in Psychological Science* 7, no. 3 (1998): s. 77-82.

351. *neleri içermesi gerektiği*: J. Levashina, C. J. Hartwell, F. P. Morgeson ve M. A. Campion, "The Structured Employment Interview: Narrative and Quantitative Review of the Research Literature", *Personnel Psychology* 67 (2014): s. 241-293.

351. *yapılandırılmış mülakatların... çok daha kestirimci olduğu*: McDaniel ve diğerleri, "Meta Analysis"; Huffcutt ve Arthur, "Hunter and Hunter (1984) Revisited"; Schmidt ve Hunter, "Validity and Utility"; ve Schmidt ve Zimmerman, "Counterintuitive Hypothesis".

352. *iş örneği testleri*: Schmidt ve Hunter, "Validity and Utility".

353. *İsrail ordusu*: Daniel Kahneman, *Think, Fast and Slow*, s. 229.

353. *pratik tavsiyeler*: Kuncel, Klieeger ve Ones, "Algorithms Beat Instinct". Bakınız ayrıca Campion, Palmer ve Campion, "Structuring Employment Interviews".

353. *"bir yanılsamanın ısrarı"*: Dana, Dawes ve Peterson, "Belief in the Unstructured Interview".

BÖLÜM 25

355. *aracı değerlendirmeler protokolü*: Daniel Kahneman, Dan Lovallo ve Olivier Sibony, "A Structured Approach to Strategic Decisions: Reducing Errors in Judgment Requires a Disciplined Process", *MIT Sloan Management Review* 60 (2019): s. 67-73.

363. *tahmin-konuşma-tahmin*: Andrew H. Van De Ven ve André Delbecq, "The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes", *Academy of Management Journal* 17, no. 4 (1974): s. 605-621. Bakınız ayrıca Bölüm 21.

KISIM 6

372. *Onların bakış açısıyla*: Kate Stith ve José A. Cabranes, *Fear of Judging: Sentencing Guidelines in the Federal Courts*, Chicago: University of Chicago Press, 1998: s. 177.

BÖLÜM 26

380. *Öncelikle, bu tür çabalar*: Albert O. Hirschman, *The Rhetoric of Reaction: Perversity, Futility,* Jeopardy, Cambridge, MA: Belknap Press, 1991.

380. *Václav Havel'den bir alıntıyla*: Stith ve Cabranes, *Fear of Judging.*

381. *"üç vuruş ve oyun dışısınız"*: Bakınız örneğin Three Strikes Basics, Stanford Law School, https://law.stanford.edu/stanford-justice-advocacy-project /three-strikes-basics/.

381. *"Woodson-Kuzey Carolina Eyaleti"*: 428 U.S. 280, (1976).

383. *peşin hüküm içerebileceğini*: Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York: Crown, 2016.

383. *"Potansiyel olarak ön yargılı"*: Will Knight, "Biased Algorithms Are Everywhere, and No One Seems to Care", *MIT Technology Review*, 12 Temmuz 2017.

384. *ProPublica*: JeffLarson, Surya Mattu, Lauren Kirchner ve Julia Angwin, "How We Analyzed the COMPAS Recidivism Algorithm", *ProPublica*, 23 Mayıs 2016; www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. Bu örnekteki ön yargı iddiası tartışmalıdır ve ön yargının farklı tanımları, zıt çıkarımlara yol açabilir. Bu örnek ve daha geniş anlamda algoritma ön yargısının tanımı ve ölçümü için, daha sonraki bir nota bakınız

384. *"tahmine dayalı polislik"*: Aaron Shapiro, "Reform Predictive Policing", *Nature* 541, no. 7638 (2017): s. 458-460.

385. *Hatta bu açıdan*: Her ne kadar bu endişe YZ'ye dayalı modeller bağlamında yeniden su yüzüne çıkıyor olsa da, sadece YZ'ye özgü değil. 1972 gibi erken bir tarihte Paul Slovic sezgiyi modellemenin var olan bilişsel ön yargıları koruyacağını ve güçlendireceğini ve hatta büyüteceğini belirtti. Paul Slovic, "Psychological Study of Human Judgment: Implications for Investment Decision Making", *Journal of Finance* 27 (1972): s. 779.

385. *tam olarak nasıl test edileceği*: Bu tartışmaya suçun tekrarlanması risk değerlendirmesi algoritması COMPAS üzerindeki ihtilaf bağlamında bir giriş için bakınız Larson ve diğerleri, "COMPAS Recidivism Algorithm"; William Dieterich ve diğerleri, "COMPAS Risk Scales: Demonstrating

Accuracy Equity and Predictive Parity", Northpointe, Inc., 8 Temmuz 2016, http://go.volarisgroup.com/rs/430-MBX-989/images/ProPublica_Commentary_Final_070616.pdf; Julia Dressel ve Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism", *Science Advances* 4, no. 1 (2018): s. 1-6; Sam Corbett-Davies ve diğerleri, "A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased Against Blacks. It's Actually Not That Clear", *Washington Post*, 17 Ekim 2016; www.washington post.com/news/monkey- cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas; Alexandra Chouldechova, "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments", *Big Data* 153 (2017): s. 5 ve Jon Kleinberg, Sendhil Mullainathan ve Manish Raghavan, "Inherent Trade-Offs in the Fair Determination of Risk Scores", Leibniz International Proceedings in Informatics, Ocak 2017.

BÖLÜM 27

390. *Yaklaşımlarının... bilebilirler*: Tom R. Tyler, Why People Obey the Law, 2. Baskı, New Haven, CT: Yale University Press, 2020.

391. *çok ünlü ve şaşırtıcı bir karar: Cleveland Bd. of Educ. v. LaFleur*, 414 U.S. 632 (1974).

392. *O dönemin nüfuzlu yorumcuları*: Laurence H. Tribe, "Structural Due Process", *Harvard Civil Rights-Civil Liberties Law Review* 10, no. 2 (Bahar 1975): s. 269.

396. *yoğun olumsuz tepkisini hatırlayın*: Stith ve Cabranes, *Fear of Judging*, s. 177.

398. *bir dizi enerjik*: Bakınız örneğin Philip K. Howard, *The Death of Common Sense: How Law Is Suffocating America*, New York: Random House, 1995 ve Philip K. Howard, *Try Common Sense: Replacing the Failed Ideologies of Right and Left*, New York: W. W. Norton & Company, 2019.

BÖLÜM 28

404. *Facebook'un Topluluk Standartları'ndan 2020*: Nefret söylemi Facebook Topluluk Standartları; www.facebook.com/communitystandards/hate_speech.

425. *New Yorker*: Andrew Marantz: "Why Facebook Can't Fix Itself", *The New Yorker*, 12 Ekim 2020.

407. *gürültülü yargıları*: Bürokratik adalet: Jerry L. Mashaw, *Bureaucratic Justice,* New Haven, CT: Yale University Press, 1983.

411. *Büyük ölçüde tam tersi*: David M. Trubek, "Max Weber on Law and the Rise of Capitalism", *Wisconsin Law Review* 720: (1972) n. 22 (aktaran Max Weber, *The Religion of China* [1951]: s. 149).

Dizin

1984 Ceza Reformu Yasası 26
360 derece puanlama sistemleri 330
"bir yanılsamanın ısrarı" 353
"Bulutlar İnekleri İyi Gösteriyor" (Simonsohn) 106
"Sihirli Sayı Yedi, Artı Eksi İki" (Miller) 209

A

ABD Ceza Komisyonu 27, 30, 85
Aboraya, Ahmed 323, 487, 488, 489
ACE-V süreci 281, 285
adalet 32, 39, 85, 86, 87, 154, 220, 226, 407, 411, 496
adli tıp 16, 96, 253, 277, 284, 285, 286, 287, 289, 290, 291, 371
adli tıp onay ön yargısı 284
ahlaki değerler 392, 394
akıcı zekâ 475
aktif olarak açık görüşlü düşünme 265, 476
alakasız bilginin etkisi 25, 253, 350
algoritma ön yargısı 383
algoritmadan kaçınma 157
alternatif yorumların bastırılması 230
Amazon Mechanical Turk 242
anlaşma yanılsaması 42, 230
anlaşmazlık 30, 33, 56, 64, 66, 172, 203, 215, 257, 316, 320, 345, 347, 414, 437
Apgar puanlaması 319, 320, 426
Apgar, Virginia 319, 320, 321, 401, 426, 487
Aracı değerlendirmeler protokolü (ADP) 367
arka kapı referansları 352
Armstong, J. Scott 295
arzu edilebilirlik ön yargısı 194
aşırı güven 247
aşırı öz güven 422
aşırı tutarlılık 198, 199, 200, 346, 426, 441, 470

Atış poligonu benzetmesi 48, 66, 186, 252
Austin, William 24
ayırt etmenin önemi 262
ayrımcılık 85, 154, 384, 385, 386, 405, 412

B

bağımsız değerlendirme 242, 353, 361
bağımsızlık 117, 349, 350, 360, 426
Baron, Jonathan 265, 476
basit kurallar (tutucu modeller) 149, 151, 156, 158, 382, 410
basit modeller 133, 137, 144
Başkan'ın Bilim ve Teknoloji Danışma Konseyi (PCAST) 287, 479
Baz oranı bilgisi 191, 200, 208, 304, 309, 360, 445
belirsizlik 35, 49, 55, 61, 90, 162, 168, 179, 230, 344, 399
Bentham, Jeremy 104, 257
Bertillon, Alphonse 280
beyaz önlük sendromu 311
Beynin işlevi 109
BIN (ön yargı, bilgi ve gürültü) tahmin modeli 305, 482
Bilgi şelaleleri 306, 363
bilgisizliğin inkârı 168
bilgiyi sıralama 253, 290, 292, 367
bilişsel ön yargı 412, 422
bilişsel tarz 265, 266
bilişsel yansıma testi (BYT) 263
Bill, cazla uğraşan muhasebeci örneği 189, 190, 191, 193, 198, 201, 202, 228, 262
bir araya toplama 306, 307, 308, 320, 349, 367, 395
bir sorunun yerine başka bir soru koymak 192, 193, 206
bireyselleştirilmiş muamele 390, 391
birinci piyangonun ürünü olarak sistem gürültüsü 96
Bock, Laszlo 349, 450, 493
boktanlığa açıklık 103
Boktanlık Üstüne (Frankfurt) 103
Breyer, Stephen 27, 453
Brier, Glenn W. 300, 301, 304
Brier skorları 300, 301
bürokratik adalet 407, 411
Büyük Beş Kişilik Özelliği modeli 138

C

Cabranes, José 28, 372, 454, 494, 495, 496
caydırıcılık 88, 374, 396
Centor puanlaması 321
Centor, Robert 321
Ceres 75
cezai hüküm 30, 253, 372, 381, 396, 434, 435
cezalandırıcı tazminat 214, 215, 216, 226, 243
Clinical Versus Statistical Prediction (Meehl) 134, 462
Corrigan, Bernard 146, 147, 464
COMPAS algoritması 384
COVID-19 krizine tepki 49
Cowgill, Bo 154, 450, 465
Curry, Stephen 94, 95

Ç

çapa atma etkisi 195
Çapalama 222, 224, 422

çapraz doğrulanmış korelasyon 146
çeşitlilik 30, 31, 37, 49, 58, 59, 86, 90, 133, 320, 428
çocuk koruma ve velayet kararları 213, 241
çoklu regresyon tekniği 133, 134, 138, 145, 146, 147, 148, 308
çoklu seçenekler ve alışverişler 64

D

Dartmouth Atlas Projesi 312, 483
davranışa dayalı değerlendirme ölçekleri 337
davranışsal olarak çapalı puanlama ölçekleri 351
değerini sorgulamak 332
değerlendiriciler arası güvenilirlik 313, 317
değerlendirme süreci 377
değerlendirmeci yargı 64, 65, 66, 67, 80, 81, 82
değerlendirmeler protokolü 16, 254, 355, 367, 383, 426, 494
Delphi yöntemi 297, 481
deontolojik etik 104
dış bakış 191, 207, 208, 253, 302, 360, 367, 375, 393, 425, 436, 445
dışlama kararı 281, 283, 284, 288, 352, 405
DNA analizi 281, 286
doğrulanabilir yargı 61, 62, 63, 416
doğrulanabilir yargıların puanlanması 416
doğrulanamaz yargılar 61, 62, 416
doğrusal regresyon modeli 133, 145
doğrusal sıralı ifşa 291
dolar olarak ceza 216, 219, 222, 225
Dreyfus, Alfred 280
Dror, Itiel 282, 283, 284, 286, 288, 289, 290, 291, 449, 478, 479, 480
durum gürültüsü kaynağı olarak ruh hâli 101, 230
durum gürültüsü kaynağı olarak stres 317
duygu temelli kısayol 194
dürtmeler 269, 270, 477
düşüncesiz tutarlılık 143
düzeltici (sonrasında) ön yargı giderme 208, 214, 269, 315, 379, 412
düzeltilmiş tahminler 447
düzenleyici davranış 400, 410
düzensiz doğrusal model (eşit ağırlık modeli) 145

E

Ebola tehdidine tepki 48
eğitim 16, 17, 53, 88, 139, 157, 171, 172, 180, 258, 259, 261, 270, 304, 306, 313, 318, 320, 344, 371, 372, 407, 411, 416, 422, 433, 465, 467, 477, 483, 486
en küçük kareler yöntemi 72, 75
ele avuca sığmaz belirsizlik 162
elverişlilik kısayolu 192
endometriozis, teşhis çeşitliliği 315
engellilik matrisi 407
erişilebilir bilginin teşhissel değerini sayısal değere dönüştürmek 443
eşit ağırlık modeli (düzensiz doğrusal model) 148, 158
eşleşen öngörüler 207, 444
eşleşen yoğunluklar 203
Eşleştirme 201, 209

Evrensel İnsan Hakları Bildirisi 411, 412
Expert Political Judgment (Tetlock) 163, 467

F

Facebook 379, 404, 405, 496
Faulds, Henry 279
faydacı hesaplama 104
Forgas, Joseph 102, 103, 458, 459, 460
Frankel, Marvin 22, 23, 24, 26, 27, 29, 30, 64, 66, 84, 85, 87, 157, 371, 380, 452, 480
Frankfurt, Harry 103, 459

G

Galton, Francis 280
Gambardi problemi 62, 203
Gates, Bill 262
Gauss, Car Friedrich 69, 72, 73, 75, 455
Gauss dağılımı 69
geciktirilmiş bütüncül yargı 349, 352, 353
geçerlilik yanılsaması 135, 140
geçici ve kalıcı faktörler 230
gelişimsel değerlendirme 330
Genel Zihinsel Yetenek (GZY) 259
Goldberg, Lewis 137, 138, 139, 140, 142, 463
GoodSell gürültüyü azaltma örneği 69, 79
Google Sırları (Bock) 349
Google'ın mülakat uygulamaları 349
Grup dinamikleri ve karar alma 111, 124
göreceli puanlamalar 335
göreceli yargılar 211, 226, 360, 367
Green Book 269, 477
grup kutuplaşması 122, 427
gürültü denetimi 15, 20, 33, 39, 69, 70, 80, 83, 84, 107, 241, 251, 282, 289, 371, 406, 433, 434, 435, 437, 438, 456
gürültünün rolü 75
gürültüye karşı 40, 225
gürültüyü azaltma 16, 30, 77, 124, 252, 253, 276, 296, 306, 324, 337, 371, 372, 374, 375, 378, 379, 380, 381, 382, 386, 387, 390, 391, 393, 394, 396, 397, 398, 399, 406, 424, 429
gürültüyü azaltmanın maliyetleri 339
gürültüyü ortadan kaldırma 16

H

hafıza performansı 108, 460
halo etkisi 331
Halpern Eleştirel Düşünme Değerlendirmesi 264, 476
Haran, Uriel 265, 476
hata denklemi 76, 80
hataların maliyeti 408
hatalı olumlular 316
hataya yatkınsa 382
hava durumu 62, 105, 106, 109, 266, 270, 296, 460
Havel, Václav 380, 495
hayat güzergâhları 176
Hertwig, Ralph 100, 458, 477
Herzog, Stefan 100, 458
Hızlı ve Yavaş Düşünme (Kahneman) 185, 353, 470
Hirschman, Albert 380, 495
hiyerarşik sınıflandırma 210

Hoffman, Paul 137, 138, 450, 490
Howard, Philip 147, 398, 464, 484, 496
Human Resources Management 330
hüküm verme kılavuzları 27, 85

I

ırksal ayrımcılık 85, 386

İ

iç ses 159, 302
içimizdeki kalabalık 99
idam cezası 381
ikinci görüş 99, 312
insan yargısına üstünlük 155, 165
İnsanlar Hakları Komitesi (Önce İnsan Hakları) 22
isabetlilik 154, 160, 225, 232, 378, 415, 424, 426
istenmeyen değişkenlik 37, 92, 384, 413
iş mülakatları 341
iş örneği testleri 494
istatistiksel düşünme 176, 180
istatistiksel ön yargı 80, 183, 185, 186, 188, 198, 200, 271, 306, 420, 438
İyi Yargı Projesi 298, 300, 301, 302

J

Jedi'nin Dönüşü (film) 193
Joan Glover-General Asistance Davası 213, 215, 216, 217, 221, 223
Jobs, Steve 262, 492
Journal of Applied Psychology 342, 466, 472, 486, 489, 490, 492, 493
Julie'nin not ortalaması örneği 204, 205, 206, 207, 208, 443, 445
jürinin karar bozması 407
jüri müzakereleri 198

K

Kadı adaleti 411
Kahana, Micheal 108, 109, 460
Kahneman, Daniel 2, 215, 353, 355, 461, 468, 469, 470, 471, 477, 478, 494
kalabalıkların bilgeliği etkisi 98
kalp hastalığı 12
kalp krizi teşhisi 166
Kant, Immanuel 104, 257
kappa istatistiği 313
kapsam duyarsızlığı 188
kapsamlı tutarlılık 229
karar gözlemcileri 274, 275
karar hijyeni 16, 252, 253, 254, 268, 276, 277, 279, 290, 291, 307, 309, 349, 355, 367, 368, 373, 379, 383, 392, 397, 409, 423, 424, 428, 438
karar toplantısı 362, 367
karmaşık yargı süreçlerinde adımlar 253, 317, 342, 349
karmaşık yargıları 253, 317, 349, 431
karşıolgusal düşünme 49
Kasdan, Lawrence 193, 194
Kefalet kararları 13
Kennedy, Edward M. 26
keyfî zulümler 23, 66
Keynes, John Maynard 300
kırık bacak prensibi 151
Kırılgan Aileler ve Çocuk Refahı Çalışması 171
kısayollar ve ön yargılar yaklaşımı 189
Kişiler arası gürültü 295, 316
kişilik benzetmesi 234
kişisel yargı 55

klinik yargı 132, 319
kolay kandırılma, ruh hali 103
Kristalize zekâ 260
kumarbaz yanılgısı 106, 460
Kuncel, Nathan 42, 142, 143, 449, 461, 463, 464, 466, 474, 475, 492, 493, 494
kurala dayalı yaklaşımlar 145
kusurlu bilgi 162

L

LaFleur, Jo Carol 391, 392, 496
Lewis, Micheal 138, 156
Lieblich, Samuel 323, 488, 489
Lucas, George 193, 194

M

Macy, Micheal 115, 461
Mashaw, Jerry 407, 496
Masumiyet Projesi 287, 479
Mayfield, Brandon 278, 279, 281, 285, 287, 289, 290
McLanahan, Sara 171, 449
medyan 36, 38, 73, 98, 122, 123, 165, 218
Meehl, Paul 134, 135, 136, 137, 140, 143, 150, 157, 168, 462, 463, 465
mekanik öngörü 133, 134, 135, 137, 155, 168
melanom 51, 316
Mellers, Barbara 164, 265, 298, 305, 449, 476, 482
Meme Görüntüleme Raporlama ve Veri Sistemi (BI-RADS) 321
meme kanseri 12, 318, 321
mini-Delphi yöntemi 297
Moneyball (Lweis) 156
Moore, Don 298, 466, 467, 472, 492
Morewedge, Carey 271, 477
Muchnik, Lev 116, 461
muğlaklık 215
mutlak puanlamalar 366
mutlak yargının sınırlamaları 208
mülakat rehberi 324
müşteriler, gürültü denetimi 19, 394
müzik indirme çalışması 112

N

naif gerçekçilik 414
Nash, Steve 94, 95
nedensel düşünme 176, 177, 181
nevi şahsına münhasır fermanlar 452
normallik vadisi 177, 178, 179, 180, 181, 246

O

Obama, Barack 46, 47, 48
Obermeyer, Ziad 166, 467
objektif bilgisizlik 129, 162, 163, 164, 166, 169, 344, 492
olay yeri izleri 280
olasılıkçı düşünme 302
OMB Circular A-4 Belge 275
onay ön yargısı 284, 285, 292, 304, 470
O'Neal, Shaquille 94, 458
O'Neil, Cathy 383, 495
orantı ölçekleri 222
ortak normlar 258
ortalama almak 296, 297, 427
ortalama hüküm süreleri 28
Ortalama Karesel Hata (OKH) 72, 74, 75, 133, 301, 415

ortalama mutlak fark 86, 87, 473
ortalamaya regresyon 444, 447

Ö

öfke hipotezi 217, 220
öfkeyi, cezai niyeti ve tazminatları karşılaştırmak 217, 218
ölçüm benzetmesi 268
ön yargı kontrol listesi 439
ön yargı kör noktası 273
ön yargı şelaleleri 286, 479
Önce İnsan Hakları (İnsanlar Hakları Komitesi) 22
öngörücü yargı 64, 81
önleyici ön yargı giderme 269

P

parçala ve fethet prensibi 425
parmak izi analizi 417
Pashler, Harold 98, 99, 100, 101, 458
patoloji, teşhis farklılığı 313
pazarlıklar 34
Pennycook, Gordon 103, 459, 475
performans değerlendirmesi 329
performans öngörüsü 42, 132, 461, 464
peşin hükümler 200, 440
Pisagor teoremi 76
piyango 35, 96, 169, 234, 420
planlama yanılgısı 186, 247, 269, 272
politik görüşler 88, 115
portföy yönetimi şirketi 38, 39
Price, Mark 94
Principles of Forecasting (Armstrong) 295, 480, 481
profesyonel doktrin 258
profesyonel yargılar 105, 226, 262, 414
proje ekibi 272, 274, 434, 436, 437, 438
proje müdürü 435, 437
ProPublica 384, 495
psikiyatri 321
psikolojik ön yargılar 185, 186, 188, 190, 200, 247, 344, 420

R

radyoloji 144, 416
Ramji-Nogales, Jaya 107, 451, 460, 473
rastgele doğrusal model 142
referans çerçevesi eğitimi 338, 339
referans sınıfı 360, 366, 425
rehabilitasyon 88, 381
Rhetoric of Reaction (Hirschman) 380, 495
Risk değerlendirme uzmanları 19, 33, 34, 35, 36, 40, 41, 42, 43, 47, 48, 51, 60, 86, 96, 97, 241, 258, 261, 282, 382, 404, 414, 415, 416, 422
Riskten kaçınma 395
Ritov, Ilana 265, 476
Robyn Dawes da Eugene 145
Romeo ve Juliet 194
Rosenzweig, Phil 247, 474
ruh hâli 48, 96, 101, 102, 103, 104, 105, 106, 110, 124, 192, 193, 199, 201, 230, 234, 292, 320, 329
ruh hâli manipülasyonu 102

S

sabit örüntü gürültüsü 232, 234, 239, 243, 244, 419, 474
Salganik, Matthew 112, 113, 115, 171, 449, 460, 461, 467
Salikhov, Marat 305, 482

Satopää, Ville 305, 306, 307, 482
satranç programı 380
Save More Tomorrow (Yarın Daha Fazla Biriktir) planları 270
saygı uyandıran uzmanlar 256, 262
Schkade, David 215, 461, 471
seçici dikkat ve seçici hafıza 58
seçme kalabalık stratejisi 297
Sellier, Anne-Laure 271, 450, 477
Sendhil, Mullainathan 151
serbest atış örneği 94, 95, 109
Sezgi 367, 426
sığınma kararları 13
sınırlı anlaşmazlık beklentisi 56, 64, 414
sıralama, performans değerlendirmesi 225, 292, 332, 334, 335, 336, 364, 367, 426
Sigorta eksperi 34
sigorta şirketinin gürültü denetimi 39
Simonsohn, Uri 106
Sistem 1 düşünme 183, 185, 263
Sistem 2 düşünme 180, 194, 208, 263
sistem gürültüsü 66, 89, 96, 186, 200, 219, 238, 241, 245, 330, 415, 417, 421, 457
sistemi manipüle etmek 394
Slovic, Paul 194, 495
sonuç ön yargıları 198, 200
sosyal etkiler 115, 117, 121
sosyal medya 379, 404
sosyal ve politik bölünmeler 403
standart sapma 70, 71, 86, 455, 457, 462
standartlar 393, 400, 401, 402, 403, 404, 406, 408, 410, 412
Statüko ön yargısı 272
Stevens, S. S. 222, 223, 470, 471
Stith, Kate 28, 372, 453, 454, 494, 495, 496
Sunstein, Cass R. 2, 215, 461, 469, 471, 477, 481
süper tahminciler 164, 256, 302, 303, 307
sürekli beta 303, 309

T

tahmine dayalı polislik 384, 495
tahmin piyasaları 297
tahmin-konuşma-tahmin 297, 363, 367
takım olma 304
tavsiye niteliğinde 29, 454
TB (tüberküloz), teşhis farklılıkları 315, 485
Tecrit 88
tekil kararlar 45, 47, 48, 254, 365
tekrarlı kararlar 47, 49
temel yükleme hatası 246
test-yeniden test güvenilirliği 97
teşhis etmek 43, 47, 323, 383, 403, 468
teşhis kararı 288, 480
teşhis koyma 53
Tetlock, Phillip 163, 164, 165, 298, 299, 300, 303, 304, 305, 306, 450, 467, 468, 482
Todorov, Alexander 242, 450, 474
toplam hata denklemi 68, 69, 72, 73, 74, 75, 76, 77, 78, 79, 80, 117, 239, 296, 416, 417
Topluluk Standartları 404, 405, 496
tutarlı keyfiyet 223, 472
tutarsızlık 66, 229, 264, 286, 294
tutumlu modeller (basit kurallar) 148
Tüberküloz 12

U

uç değerler 447
Ulusal Basketbol Birliği (NBA) 94
Uygulama Standartları 405

Ü

üç vuruş politikası 381
üst geçit problemi 459

V

vaka ölçekleri 338, 339
vakaya göre yargıç 90
Venedik Taciri (Shakespeare) 389
Vul, Edward 98, 99, 100, 101, 458

W

Wainer, Howard 147, 464
Weapons of Math Destruction (O'Neil) 383, 495
Weber, Max 411, 496
Welch, Jack 332
Williams, Thomas 24, 413, 452
Woodson-Kuzey Carolina Eyaleti davası 381

Y

Yang, Crystal 29, 454
yapay öğrenme modelleri 145, 151
yapılandırılmış davranışsal mülakatlar 351, 494
yapılandırılmış yargı 353
yaratıcılık 57, 237, 260, 397, 424
yargıcın modeli 142, 232
yargıçlar arası uyuşmazlıklar 28
yargılama korkusu 372
yargının tamamlandığına dair içsel sinyal 61, 160
yargısal takdir yetkisi 31
yatay gürültü 89, 91, 93, 153, 154, 219, 240, 241, 244, 245, 326, 328, 340, 418, 438, 457, 489
yatay hatalar 87
Yetişkin Karar Alma Kabiliyeti ölçeği 264
yoğunluk ölçekleri 203
Yu, Martin 142, 143, 461, 463, 464
yüzdelik uyum (YU) 161
YZ (Yapay Zekâ) 144, 150, 151, 152, 156, 158, 166, 173, 252, 318, 487, 495

Z

zekâ 133, 144, 166, 197, 259, 260, 266, 302, 318, 326, 421, 423, 475
Zihinsel Bozuklukların Tanısal ve Sayımsal El Kitabı, 3. Baskı (DSMIII) rehberler 322
Zihinsel Bozuklukların Tanısal ve Sayımsal El Kitabı 4. Baskı (DSM-IV) rehberler 322
Zihinsel Bozuklukların Tanısal ve Sayımsal El Kitabı 5. Baskı (DSM-5) rehberler 322
zorunlu idam cezası 381
zorunlu sıralama sistemi 332, 334
Zuckerberg, Mark 262



DANIEL KAHNEMAN, 2002 yılında Nobel Ekonomi Ödülü'nü ve 2013 yılında da Ulusal Özgürlük Madalyası'nı kazandı. Princeton Üniversitesi'nde Eugene Higgins Psikoloji Profesörü ve Princeton Kamu ve Uluslararası İlişkiler Okulu'nda da Halkla İlişkiler Profesörüdür. Aralarında Amerikan Psikoloji Derneği'nden Psikolojiye Ömür Boyu Katkı Ödülü'nün de bulunduğu çok sayıda ödülün sahibi olan Kahneman, uluslararası çoksatan *Hızlı ve Yavaş Düşünme* kitabının da yazarıdır.

OLIVIER SIBONY, HEC Paris'te Strateji ve İş Politikası Profesörüdür. Daha önce kıdemli ortak olduğu McKinsey & Company şirketinin Paris ve New York ofislerinde yirmi beş yıl çalıştı. Sibony'nin stratejik karar vermenin kalitesini artırmaya yönelik araştırması, *Harvard Business Review* ve *MIT Sloan Management Review* dâhil olmak üzere birçok yayında yer aldı.

CASS R. SUNSTEIN, *New York Times*'ın çoksatan *The World According to Star Wars* ve (Richard H. Thaler ile birlikte yazdığı) *Dürtme* kitaplarının yazarıdır. Davranışsal Ekonomi ve Kamu Politikası Programı'nın kurucusu ve yöneticisi olduğu Harvard Robert Walmsley Üniversitesi'nde profesördür. 2009 yılından 2012 yılına kadar Beyaz Saray Bilgi ve Düzenleme İşleri Ofisi'nde yöneticilik yapmış ve 2013 yılından 2014 yılına kadar Başkan Obama'nın İstihbarat ve İletişim Teknolojileri İnceleme Grubu'nda görev almıştır.

Kapak tasarımı: Julian Humphries
Kapak fotoğrafı: © Shutterstock
Yazar fotoğrafları, yazarların izniyle yayımlanmıştır.
Kapak: © 2021 Hachette Book Group, Inc.

NEREDE İNSAN YARGISI VARSA ORADA GÜRÜLTÜ VARDIR...

Aynı şehirdeki iki doktorun aynı hastalara farklı teşhisler koyduğunu ya da aynı mahkemedeki iki hâkimin aynı suçları işleyen kişilere farklı cezalar verdiğini düşünün. Şimdi aynı doktorun ve aynı hâkimin, sabah veya öğleden sonra ya da çarşamba yerine pazartesi olmasına yahut henüz öğle yemeği yememesine bağlı olarak farklı kararlar verdiğini hayal edin. İşte bu aynı olması gereken yargılardaki değişkenlik, gürültü örnekleridir.

Gürültü kitabında Daniel Kahneman, Olivier Sibony ve Cass R. Sunstein; tıp, hukuk, halk sağlığı, ekonomik tahmin, adli tıp, çocuk koruma, yaratıcı strateji, performans incelemesi ve işe alma dâhil olmak üzere birçok alanda gürültünün nasıl hatalar ürettiğini gösteriyor. Her ne kadar insanların yargıda bulunduğu ve karar verdiği her yerde gürültü bulunabilse de, bireyler ve kuruluşlar benzer şekilde gürültünün etkisini görmezden gelirler.

Yeni fikirlerle dolu *Gürültü*, insanların karar verme sürecinde gürültüye ve ön yargıya nasıl ve neden bu kadar duyarlı olduğunu açıklıyor. Hepimiz düşündüğümüzden daha fazla kötü kararlar veririz. Birkaç basit çözüm sunan bu çığır açıcı kitap, nasıl daha iyi yargıda bulunabileceğimizi ortaya koyuyor.

"Anıtsal bir eser. *Gürültü*, son on yıl içinde okuduğum en önemli kitap olabilir. Gerçekten yeni bir fikir ortaya koyuyor. Laf aramızda bu o kadar önemli bir fikir ki hemen uygulamaya koyacaksınız."

Prof. Dr. Angela Duckworth, *Azim* Kitabının Yazarı





www.pegasusyayinlari.com
ISBN: 978-625-410-389-6