

فریب و ریاکاری با داده‌ها: کلاهبرداری در علوم*

دیوید هند

ترجمه حمید پزشکی

دو سال قبل جامعه علمی از شنیدن خبر متهم شدن «وو ساک هانگ^۱» محقق کلونینگ و سلول‌های بنیادی به جعل نتایج کارش، بسیار حیرت زده شد. نه فقط او، که زیگموند فروید^۲، نیوتن^۳، و برخی از برندگان جایزه نوبل هم متهم به تقلب در داده‌ها شده‌اند. دیوید هند^۴ با بررسی مسأله تقلب در علوم درمی‌یابد که این کارچندان هم نادر نیست.

وقتی «وو ساک هانگ» دانشمند بیومدیکال کره‌ای ادعا کرد توانسته با موفقیت شبیه‌سازی گاوها و یک سگ را به انجام برساند، توجه مطبوعات و مراکز علمی را به خود جلب نمود. این توجه با پیشرفت‌های مهم دیگری در مورد تحقیق سلول‌های بنیادی در آزمایشگاه او ادامه یافت. ([#-note-16](http://en.wikipedia.org/wiki/Hwang-Woo-Suk))

ولی به زودی این توجهات تغییر جهت دادند زیرا همکار آمریکایی او جرالد شتن^۵ به دلیل آن که از روند اهدای تخمک انسانی راضی نبود از ادامه تحقیقات با هانگ امتناع ورزید. بررسی‌ها نشان داد که او یا شرکای نزدیکش به زنان برای اهدای تخمکشان پول داده‌اند. با بالا گرفتن موضوعات اخلاقی، مسائل سرعت گرفتند. در ماه مارس ۲۰۰۶ او از شغلش در دانشگاه بین‌المللی ستول اخراج شد و در ماه می آن سال متهم به کلاهبرداری گشت. اعضای گروه تحقیق دریافتند هانگ، بسیاری از داده‌ها در مورد سلول‌های بنیادی را خود به وجود آورده و متهم شد که ۳ میلیون دلار از ۴۰ میلیون دلار پول تحقیق خود را به طور نادرست برای چیزهایی مثل خرید ماشین برای

(* این نوشته ترجمه‌ای از مقاله زیر است:

Hand, D; (2007) Deception and Dishonesty with Data: Fraud in Science. Significance, issue 1, Vol. 4, pp 22-25

1) Woo Suk Hwang 2) Sigmund Freud 3) Newton 4) David Hand 5) Gerald Schatten

همسرش و هدیه برای سیاستمداران مصرف کرده است. او اکنون منتظر دادگاه است و احتمالاً به زندان محکوم خواهد شد. گاهی مردم یک تصور ترجیحاً ایده‌آل از دانشمندان دارند. آنان اغلب به عنوان مشاهده‌گرانی بی‌احساس از طبیعت دیده می‌شوند که بدون آشفته‌شدن از انگیزه‌های متناقض، اسرار را بیرون می‌آورند و قادرند آنچه را که موجب ناراحتی ما می‌شود، هدایت کنند. به هر حال حقیقت، همان طور که هر کسی در دنیای علوم کار کند می‌داند، این است که دانشمندان هم انسان هستند و مثل بقیه در معرض تأثیرات مشابه قرار دارند. محیط‌های علمی هم محیط‌هایی انسانی هستند و دانشمندان مثل هر کس دیگری، رؤسایی با درخواست‌های غیرممکن دارند، دارای دستیارانی هستند که آنچه را از آنها خواسته شده انجام نمی‌دهند و با آن شرایط داورانی متخصص هم درباره آنها قضاوت می‌کنند. مخصوصاً دنیای علم همانقدر به شدت رقابتی است که دنیای تجارت و بازرگانی.

رقابت، از راه‌های متعددی آشکار می‌شود. پیشرفت شغلی به مقالات موفق فرد بستگی دارد و لااقل در زمینه‌های دانشگاهی به جوایز نقدی. موفقیت، شناخت گروه متخصص، شاید قدرت، گاهی حتی وعده ثروتمندان را که با یک امتیاز به موقع پشتیبانی می‌شود، ایجاب می‌کند. حتی می‌تواند نوع اشتها را تعیین نماید که از انتساب یک اکتشاف یا پیشرفت به یک شخص حاصل می‌شود. تعجبی ندارد که بعضی محققین با توسل به کلاهبرداری می‌خواهند مطمئن شوند که موفق هستند. برای هانگ جوایز و فشارها شدیدتر از خیلی‌های دیگر بود. به او قبل از سقوطش، عنوان رسمی اولین دانشمند بی‌نظیر کره‌ای داده شد و یک سری تمبر به افتخار او چاپ شده بود.

در مقیاس کوچک‌تر، دیگران نیز تسلیم وسوسه شده‌اند. مارتینسون و دیگران در مقاله‌ای در Nature دریافتند که از ۳۲۴۷ دانشمند بررسی شده به طور ناشناس، ۳/۵٪ به داشتن داده‌های دروغین در کار تحقیقی خود اعتراف کردند. ۶٪ هم اعتراف کردند که اگر داده‌ای با تحقیق قبلی‌شان ضدیت داشته، آن را ارائه نکرده‌اند. دانشمندان در اواسط دوران کاری خود، احتمالاً بیش از جوانترها در این رفتارهای نادرست زیاده‌روی کرده‌اند که شاید به دلیل وجود فشار بیشتر روی آنان در این مرحله از دوران کار باشد.

هرچند در بعضی از کلاهبرداری‌ها مثل کلاهبرداری بانکی، مرتکبین از همان ابتدا غیرصادق هستند و عمداً در معرض فریب خوردن قرار می‌گیرند، من فکر می‌کنم بیشتر دانشمندان متقلب به تدریج خود را اسیر کلاهبرداری می‌بینند. برای فهمیدن دلیل این کار، به داده‌هایی که دانشمندان جمع می‌کنند فکر کنید.

داده‌ها، حتی قبل از ارائه به یک آماردان برای تحلیل، اغلب خود نتیجه یک فرایند انتخاب هستند. محقق که در زمان اندازه‌گیری، یک ابزار یا نمونه را کنار می‌گذارد ممکن است احساس کند که اندازه‌گیری مربوطه باید حذف شود. اگر آنها متوجه شوند که یکی از کنترل‌کننده‌ها از مقدار خواسته شده متفاوت می‌باشد، ممکن است احساس کنند که نتیجه نباید در تحلیل وارد شود. در واقع اگر متوجه شوند که مقدار پاسخ متغیر از یکی از اندازه‌گیری‌های تصادفی کاملاً با دیگران متفاوت

است، حتی اگر هیچ توضیحی برای آن نداشته باشند، ممکن است شک کنند و آن را از آنالیز حذف نمایند. هیچ قانونی در مورد تعیین حد و حدود این کار وجود ندارد. به ناچار مسائل معنوی وارد جریان می‌شوند و با تمرین آزادی بیشتر در انتخاب مقادیرهایی که فرد می‌خواهد قبول کند، می‌تواند هر نتیجه‌ای که بخواهد به دست آورد و از هر فرضیه‌ای که بخواهد پشتیبانی کند.

رابرت میلیکان، برنده جایزه نوبل فیزیک سال ۱۹۲۳، به سبب اندازه‌گیری بار الکترون در قطره روغن شهرت دارد. آزمایش او یک پیروزی بزرگ است و نشان می‌دهد که الکتروسیسته یک جریان نیست بلکه به صورت واحدهای شبه ذره‌ای جداگانه در حرکت است. بار الکتریکی می‌تواند فقط در مضرب‌های عدد کامل واحد پایه‌اش وجود داشته باشد و شکستن بارهای الکتریکی امکان ندارد. ولی با وجود این، موضوع خیلی بحث برانگیز بود. میلیکان (که در اواسط دوران کاری‌اش بود و چندان مشهور نبود) نیاز داشت توجه دیگران را به این موضوع جلب کند. پس به نفع میلیکان بود که داده‌اش، از نظریه پشتیبانی کند و بسیاری از مشاهدات منتشر شده او نیز این کار را می‌کردند. به هرحال، همه اندازه‌گیری‌های ثبت شده وی این طور نبودند. مثلاً در مقاله‌ای در *Physical Review* گزارش داد که «آزمایش او نه در مورد یک گروه انتخاب شده از قطره‌ها بلکه حاصل کار بر روی همه قطره‌ها در ۶۰ روز متوالی است.» ولی او حقیقت را نمی‌گفت. بررسی دفترچه‌های یادداشت اصلی او توسط رابرت هولتون در سال ۱۹۷۸ نشان داد بسیاری از مشاهدات ثبت شده وی در بین آن آنالیزها وجود ندارند. میلیکان کار بر روی ۵۸ قطره کوچک را گزارش نمود ولی در واقع ۱۷۵ قطره ثبت کرد. یک نظریه خوشبینانه این خواهد بود که او ۱۱۷ نتیجه را حذف کرده است،

(<http://www.sigmaxi.org/meeting/archive/forum.2000.millikan.shtml>)

نه به دلیل آن که این نتایج بر ضد فرضیه‌اش بودند بلکه به دلیل پایین بودن کیفیت داده‌ها این کار را کرد، با این که در دفترچه یادداشت خودش نوشته بود «این تقریباً کاملاً درست است»، «این نتیجه قشنگ را منتشر کن»، «خیلی پائین، چیزی غلط است»، «عدم توافق، درست در نمی‌آید.» این که چقدر جدی چنین چیزهایی مورد توجه قرار می‌گیرند، اغلب به این بستگی دارد که بعداً تحقیق از نظریه پشتیبانی کند. بدبختانه برای دانشمندی که می‌خواهد داده‌ها را به نظریه مرتبط کند علم یک فرایند خودکنترل است. نظریات غیرصحیح می‌توانند مدتی دوام بیاورند ولی بالاخره در مقابل مدارک مخالف خود، تسلیم می‌گردند.

سیر رویدادها برای «پیل‌داون من»^۱ نیز به همین صورت بود. در سال ۱۹۱۲ در ساینس استخوان فک و جمجمه‌ای توسط «چارلز داوسن»^۲ به دست آمد که ادعا می‌کرد (و در آغاز مورد قبول هم واقع شد) که فسیل حلقه گمشده کمیاب بین انسان و میمون است. ولی بعد کنار گذاشته شد چون ثابت شد جعلی است و حتی قبل از ۱۹۵۳ کاملاً فراموش گردید. جمجمه فقط ۶۰۰ سال عمر داشت و استخوان فک متعلق به اورانگوتان بود. هنوز هم کاملاً معلوم نشده که آیا داوسن فسیل را اول دفن و سپس بیرون آورده است یا نه. همچنین در سال ۱۹۲۰ «مورد قورباغه قابل»^۳

1) Piltdown Man 2) Charles Dawson 3) Case of the Midwife Toad

می‌خواست از نظریه لامارکی یعنی توارث خصوصیات اکتسابی بر ضد داروینسم پشتیبانی کند. ولی نوارهایی که ادعا می‌شد نتیجه کلیدی می‌باشند، تقلبی از آب درآمدند. «کامرر» که آزمایش را انجام داد خودکشی کرد. بنابراین کلاهبرداری علمی یک پدیده جدید نیست. در واقع به قدر کافی باعث نگرانی چارلز بابیج^۱، یکی از مؤسسين انجمن سلطنتی آمار بوده که اقدام به ایجاد یک طبقه‌بندی کند: «انواع متعددی از فریب وجود دارند که در علوم به کار رفته‌اند ولی تعداد کمی شناخته شده‌اند مگر آنها که جدیدتر هستند... این‌ها ممکن است تحت عنوان کلک زدن، جعل کردن، آرایش کردن و پختن طبقه‌بندی شوند.» او کلک زدن را به عنوان «قصید لو رفتن در آخر کار» توصیف کرد تا کسانی را که دنبال آن می‌روند مسخره کرده باشد، ولی جعل کردن را قصد دوام یافتن تا مدتی طولانی معنی کرد. آرایش کردن به معنای «کوتاه کردن مقداری از مشاهداتی که از حد متوسط، افراط کرده‌اند و چسباندن آنها به مشاهداتی که زیادی مختصرند» می‌باشد؛ و بالاخره پختن را به عنوان «انتخاب داده‌ای که از بحث فرد پشتیبانی کند» مشخص کرد، اتهامی که علیه میلیکان وارد شده است از این نوع می‌باشد.

البته همه دانشمندانی که داده‌ها را با انتخاب تغییر شکل می‌دهند، از کارشان آگاه نیستند. آنان ممکن است از پدیده آماری مثل آریبی انتخاب بی‌اطلاع باشند و شاید همه خدمات و دوره‌های آموزشی کمکی در مورد آمار باید شامل توضیحاتی در مورد میزان جدی بودن آن باشد. یا شاید درک نکنند چه اتفاقی در حال شکل‌گیری است. مثل موردی با کشف قلبی رابطه بین خیال پردازی و واکسن سرخک، اوربون و سرخچه با آنالیز ابتدایی براساس آزمایش بر روی یک گروه بسیار انتخابی کودکان. چنین مواردی نتیجه کلاهبرداری نخواهد بود بلکه صرفاً نتیجه جهل آمار پایه‌ای هستند.

اگر انتخاب داده‌ها می‌تواند از یک طرف بدون جهت گیری باشد و از طرف دیگر عمداً دستکاری شده باشد، پس قادر است خود اعداد را تغییر دهد. دستکاری صریح یعنی وقتی که فرد می‌گوید چه کاری انجام داده، کاملاً قابل قبول است. بنابراین دیگران می‌توانند تصمیم بگیرند که آیا به نظر آنها نتایج، تغییر شکل داده شده‌اند یا نه. ولی دستکاری پنهانی یعنی وقتی فرد تغییر را پنهان می‌کند، چیز کاملاً متفاوتی است. هر چه ابزارهای نرم‌افزاری پیشرفت می‌کنند، این کار آسان‌تر می‌شود. وسوسه تصحیح سطحی فقط چند مورد از مشاهدات برای این که P مقدار را بامعنا کنند تا مقاله منتشر شود، باید گاهی بسیار قوی باشد.

مشکل این است که علوم در خلأ ساخته نشده‌اند. داده‌ها فقط می‌توانند در زمینه یک نظریه تفسیر شوند. بنابراین طبیعتاً باید تمایلی برای حذف کردن مقدارهایی که با نظریه جور نیستند وجود داشته باشد. میلیکان ممکن است دلایل قابل توجیهی برای انتخاب مشاهدات خاصی داشته ولی تصور کردن دیگران در حال انتخاب یا اصلاح داده‌ها با حداقل بی‌گناهی آسان نیست. اعتماد به نفس کافی یا جهل می‌تواند به آسانی به این نظریه منتهی شود که من می‌دانم نظریه‌ام

1) Charles Babbage

درست است و من در درست کردن داده‌ها، برحق می‌باشم تا اشتباهات را برطرف کنم و کاری کنم نظریه‌ام قطعی‌تر به نظر آید. میلیکان فقط یک نمونه است. بررسی دقیق دفترچه یادداشت‌های حتی بزرگترین دانشمندان گذشته ثابت می‌کند که خیلی از آنان این کار را کرده‌اند. نیوتن، مندل، پاستور، گاليله و کپلر همه، ابرهای تردید در بالای سر خود داشته‌اند. شواهد قطعی نشان می‌دهند که زیگموند فروید در ادعاهایش به گمراهی رفته است.

البته، فرد می‌تواند از حد فقط انتخاب و اصلاح هم فراتر رود. در نهایت اگر داده‌ایی که با زحمت جمع‌آوری شده نظریهٔ فرد را تأیید نکند، با عوض شدن، می‌تواند اصلاح شود. در سال ۲۰۰۵، ادارهٔ بهداشت عمومی آمریکا^۱ دریافت که استیون لیدون^۲ با تحریف نمونه‌های DNA و ساختن ترکیبات دروغین برای آزمایشات انجام شده در آزمایشگاهش به گمراهی افتاده تا از یافته‌های ناقص ادعایی خود در فرایند بازسازی DNA پشتیبانی کند.

(<http://ori.dhhs.gov/miscoduct/cases/Leadon.shtml>.)

حتی یک گام بلندتر هم می‌توان برداشت. وقتی کسی آنقدر در این راه لغزنده پایین رفته که داده‌ها را خود انتخاب و مقدارها را برای پشتیبانی از نظریه‌اش اصلاح یا تحریف کند، کافی است یک گام کوچک بردارد تا بفهمد چیزها آسان‌تر می‌بودند اگر از همان ابتدا داده‌ها را خود ابداع می‌کرد. در زمینه‌ای دیگر، مورد «سنگ جدول پیاده‌رو» برای توصیف سرشماری کننده‌هایی به کار می‌رود که روی سنگ جدول پیاده‌رو می‌نشینند و شمارش می‌کنند.

این، همان مورد «جان دارسی^۳» است. او یک متخصص داروسازی برای کمک به معالجهٔ سکتة قلبی در هاروارد بود، که عده‌ای از همکارانش به او مشکوک شدند. سرپرست آزمایشگاه از او در مورد داده‌هایش پرس و جو کرد تا نتایج را بررسی کند ولی دارسی به جای ارائهٔ داده‌ها، شروع به اندازه‌گیری‌های جدید کرد، او همه اندازه‌گیری‌ها را در یک روز انجام داد ولی در روزهای مختلفی ثبت نمود. بررسی بعدی نشان داد داده‌های مقالات دیگر حتی زمانی که او هنوز دانشجوی دوره کارشناسی بود نیز دستکاری و یا اختراع شده بودند.

جان هندریک شون^۴ نمونهٔ جدید دیگری در زمینهٔ فیزیک می‌باشد. او نیز مانند دارسی به نظر می‌آمد یک محقق خلاق در سال ۲۰۰۱ باشد که هر ۸ روز یک مقاله مشترک می‌نوشت. ولی دیگر محققین در نتایج کار او متوجه کارهای خلاف آشکاری شدند.

مخصوصاً که این نتایج بیش از حد درست به نظر می‌رسیدند. بررسی‌ها آشکار کردند که دو آزمایش متفاوت، مقدارهای مشابهی به دست آورده بودند و بررسی دقیق‌تر موارد دیگری را نشان داد که همان داده‌ها در آزمایشات دیگری گزارش شده بودند. معلوم شد که سری داده‌های دیگر به طور تجربی جمع‌آوری نشده ولی توسط کامپیوتر و با فرایندهای ریاضی حاصل آمده است.

کلاهداری علمی عواقبی دارد. در یک سطح دیگر، در کار یک کلاهدار پیشرفتی دیده می‌شود

1) US Public Health Service 2) Steven Leadon 3) John Darsee 4) Jan Hendrik Schön

که شاید استحقاق آن را نداشته باشد. ولی گرفتاری‌های بسیار جدی‌تری وجود دارند. کلاهبرداری در تحقیقات پزشکی باعث بروز ضایعات مستقیم در مردمی می‌شود که از داروهای نامناسب استفاده می‌کنند. یک مقاله لیدون که در سال ۱۹۹۷ منتشر شده ۲۲۷ بار مورد ارجاع داشت تا این که بالأخره ملغی شد. این موضوع نشان می‌دهد یک کلاهبرداری اگر کشف نشود چه اثر مخربی می‌تواند داشته باشد.

همچنین تأثیر بر علوم و تصور خود علوم نیز وجود دارد. تبلیغات در مورد دانشمندان کلاهبردار می‌تواند از خود کلاهبرداری اثر مخرب بیشتری داشته باشد (که به همین دلیل، اغلب پنهان می‌شود - همان‌طور که یک بانکدار زمانی به من گفت: «بانک من هیچ نوع کلاهبرداری ندارد.») که همین، آن را منحصر به فرد می‌سازد. در جنگی که علم در مواجهه با ادعاهای بی‌معنای شبه‌علمی دارد آخرین چیزی که نیاز دارد موارد عمومی شده‌ای است که به این اشاره می‌کند که قهرمانان آن قابل اعتماد نیستند. از طرف دیگر، ممکن است وقتی یک کلاهبرداری کشف می‌شود، فریادهای ناشی از تعجب و شلوغ کردن زیاد، این عقیده را تقویت می‌کند که این امر خیلی بعید است.

ولی چقدر بعید است؟ کلاهبرداری چقدر در علوم شایع است؟ ماهیت کلاهبرداری جواب دادن به این پرسش را بسیار مشکل می‌سازد. حداقل یک تعریف دقیق از معنای کلاهبرداری را می‌طلبید که با چند مدل‌سازی آماری پیچیده پشتیبانی شود به طوری که با دیدن نوک کوه یخ، اندازه آن را تخمین بزنند. چند مورد اخیر با این توافق خاتمه یافته‌اند که به شرط عدم پذیرش مسؤلیت تحقیقات علمی دیگر، هیچ عملی علیه فرد دانشمند صورت نگیرد. (در تمام مواردی که من از آنها آگاه شده‌ام، فرد دانشمند مذکور بوده است.) این ثابت می‌کند که خیلی از موارد اصلاً گزارش نمی‌شوند.

مقاله‌ای در Nature تعداد دانشمندانی که مرتکب جعل، تحریف یا سرقت ادبی می‌شوند را بین ۱٪ تا ۲٪ تخمین زده است.

سؤال این است که چگونه تقلب علمی را آشکار کنیم. یک ویژگی مشترک بین بسیاری از موارد کشف شده این است که اول یک خلاف کوچک کشف می‌شود و بررسی دقیق‌تر، خلاف بزرگتری را آشکار می‌کند که هرچه آزمایشات ادامه می‌یابند عمیق‌تر می‌شوند تا این که بالأخره تخلفات فاصله‌دار بزرگی معلوم می‌شوند. بعضی وقت‌ها تاریخچه‌ای از ابزارهای آماری دقیق وجود دارد که برای آشکارسازی کلاهبرداری به کار می‌روند. مانند مورد جان هندریک شون (و نیز نیوتن و مندل) که در آنها داده‌ها بیش از آن خوب بودند که درست باشند. آنها بیش از حد با فرضیه‌های تجربی منطبق بودند. گاهی داده‌ها به قدر کافی تصادفی نیستند (انسان‌ها در ساختن اعداد تصادفی خیلی خوب کار نمی‌کنند) و آزمایش‌های آماری این را آشکار می‌کنند. مثلاً توزیع بنفورد^۱، توزیعی را به دست می‌دهد که برای اولین رقم در مجموعه‌های مشاهدات طبیعی تحت شرایط خاص انتظار خواهیم داشت و خروج از این توزیع می‌تواند ثابت کند اعداد تصنعی‌اند.

1) Benford

انسان چه باید بکند اگر به کلاهبرداری مشکوک شد؟ این به وضوح مشکل ساز است اگر یک محقق پایین‌تر به سرپرست خود مشکوک شود. برای یک کلاهبردار رده بالا غریب نیست که اگر متهم شد آن را پاک کند. ووساک هانگ، دیگران را در تهیه داده‌های غلط مقصر می‌داند. برعکس آن نیز دشوار است. مقالات توضیح می‌دهند داده‌های تحریف شده جان داری، با همکاری افراد پایین‌تر از او نوشته شده‌اند. آیا آنها باید برای عدم تشخیص کار خلاف جریمه شوند؟ تا چه حد می‌توان در کار دیگران دقت کرد؟ بالاخره جایی باید اعتماد وجود داشته باشد.

افراد و ادارات مختلفی برای کنترل کلاهبرداری علمی به کار گرفته شده‌اند.

US Department of Health and Human Services اداره‌ای در مورد درستی تحقیقات دارد، که در سال ۱۹۹۲ به عنوان ترکیبی از دو سازمان تأسیس شد و در سال ۱۹۹۳ از سازمان‌های مالی مستقل شد. در بریتانیا Committee on Publication Ethics سازمانی برای ناشرین مجلات علمی است که مسائل اخلاقی در تحقیقات را بررسی می‌کند.

بدبختانه با وجود چنین سازمان‌هایی از آنجا که علم یک حرفه انسانی است مهم نیست با چه دقتی مراقبت و کنترل بشود، کلاهبرداری هرگز کاملاً ناپدید نمی‌شود. مانند فقر، کلاهبرداری همیشه با ما خواهد بود.

مراجع

- [1] Martinson, B. C., Anderson, M.S. and de Vries, R. (2005) Scientists behaving badly. *Nature*, 435, 737-738.
- [2] Koestler, A. (1973) *The Case of the Midwife Toad*. New York: Random House.

مترجم: حمید پزشک

دانشکده ریاضی، آمار و علوم کامپیوتر، قطب بیومتمتیک پردیس علوم، دانشگاه تهران

pezeshk@khayam.ut.ac.ir