Blind Assessment in Turn-of-the-Century Physiology and Pharmacology

The hypnotism-suggestion debate generated enormous scientific and public awareness (at least on the European continent) of the potential power of suggestion. This recognition gradually influenced and promoted a perception of the necessity to adopt blind assessment with placebo controls when investigating the effects of stimulants and other substances on human beings.

Brown-Séquard's Testicular Extract

The trail of blind assessment from neurology and psychiatry to physiology and pharmacology, as best I can uncover, began on 1 June 1889. Charles Édouard Brown-Séquard (1817–94), one of the most prominent scientists of his time, stunned the Société de Biologie in Paris by announcing that subcutaneous injections of animal testicles could rejuvenate physical and mental health. Brown-Séquard ended his original announcement with an unusual caveat for a research physiologist: with the Charcot-Bernheim furor in the background, he was forced to remark that he did not want to discuss to what extent "autosuggestion, without hypnosis" (*une sorte d'auto-suggestion, sans hypnotisation*) could explain his outcome, and he dismissed the possibility as remote.[83] Despite all the attention to blind assessment in the hypnotism-suggestion literature, Brown-Séquard followed the "mainstream" convention of his science and did not consider blind assessment helpful in physiology or medicine. Rather, he advocated the old system of reliance on experienced investigators and called on other "physiologists advanced in age" to perform similar self-experiments to confirm his results with case histories.[84]

---

accompany valerian followed, and then he found that he had made a mistake and was holding the bottle containing valerian" (Arthur Thomas Myers, "On the Action of Drugs at a Distance," *J. Soc. Psychical Res.*, 1885, *2:* 58–62, quotation on p. 61).

A commission of the Academy of Medicine headed by Georges Dujardin-Beaumetz replicated versions of these experiments in what seems an intentional double-blind manner. Tubes with genuine drugs were matched with identical empty tubes; the tubes were numbered and then covered with paper so that the experimenter could not tell whether substance or sham was being tested: Georges Dujardin-Beaumetz, "Sur l'action des médicaments à distance," *Bulletin Général de Thérapeutique Médicale et Chirurgicale,* 1888, *114:* 241–61. Most of the other independent replications seem to have been single-blind; e.g., J. Voisin, "Suggestion, auto-suggestion et vivacité du souvenir dans le sommeil hypnotique.—Action des médicaments à distance," *Revue de l'Hypnotisme,* 1888, *2:* 209–11.

83. Charles Édouard Brown-Séquard, "Des effets produits chez l'homme par des injections sous-cutanées d'un liquide retiré des testicules frais de cobaye et de chien," *Comptes Rendus de la Société de Biologie,* 1889, *41:* 419.

84. Ibid.

For most of the medical and scientific profession the debate on the extract was framed with the need to produce verified case histories.[85] But the fear of delusion began to be raised, especially as testicular extract became an international popular sensation.[86] Segments of professional medicine even began to raise the likelihood of "suggestion and the influence of the imagination."[87] A handful of French researchers considered blind assessment. Less than two months after Brown-Séquard's original announcement, Dr. M. G. Variot at the Hôtel Dieu performed the first independent confirmation of the new therapy: he gave three elderly men the injections, telling them that they were receiving "fortifying" (*fortifiante*) injections, and obtained results identical to Brown-Séquard's.[88] Subsequently, to address the suspicion of suggestion, Variot gave two other patients injections of water with the same instruction, with no effect; these two individuals were later injected with the extract and underwent youthful changes.[89] These two patients and the blind test of the extract are, to my knowledge, the first instance of a substance advo-

85. For example, the *Brit. Med. J.*'s report of Brown-Séquard's observations stated that "they [Brown-Séquard's observations] would require to be rigidly tested and fully confirmed by other self-experimenters before they were likely to meet with general acceptance" (*Brit. Med. J.*, 1889, *1:* 1416); no mention is made of the need to make blind assessments. Throughout the debate, Brown-Séquard's defense comprised long recitals of the effects of the extracts on hundreds of cases of such diseases as cancer and tuberculosis. See Charles Édouard Brown-Séquard, "On a New Therapeutic Method Consisting in the Use of Organic Liquids Extracted from Glands and other Organs," *Brit. Med. J.*, 1893, *1:* 1212–14. He never seemed to discuss blind assessment, with the exception of the one passing reference discussed below (see n. 89).

86. The popular press spoke of the "fountain of perpetual youth," "elixir of youth," the "Alchemist's Dream," and said that a "cult of injection spread like wildfire" (James M. D. Olmsted, *Charles-Édouard Brown-Séquard: A Nineteenth-Century Neurologist and Endocrinologist* [Baltimore: Johns Hopkins Press, 1946], pp. 210–11). Also see Michael J. Aminoff, *Brown-Séquard: A Visionary of Science* (New York: Raven Press, 1993); Merriley Borell, "Brown-Séquard's Organotherapy and Its Appearance in America at the End of the Nineteenth Century," *Bull. Hist. Med.*, 1976, *50:* 309–20.

87. Editorial, "Animal Extracts as Therapeutic Agents," *Brit. Med. J.*, 1893, *1:* 1279.

88. M. G. Variot, "Trois expériences sur l'action physiologique du suc testiculaire injecté sous la peau, suivant la méthode de M. Brown-Séquard," *Comptes Rendus de la Société de Biologie,* 1889, *41:* 451–54.

89. Brown-Séquard reported on the Variot experiments and the subsequent single episode of blind assessment in Charles Édouard Brown-Séquard, "Remarques à l'occasion du travail de M. Variot, sur les injections de liquide testiculaire chez l'homme," *Comptes Rendus de la Société de Biologie,* 1889, *41:* 454–55, quotation on p. 455. (Also see idem, "The Effects Produced on Man by Subcutaneous Injections of a Liquid Obtained from the Testicules of Animals," *Lancet,* 1889, *2:* 105–7, which also provides an English summary of Brown-Séquard's original French report.) This is the only mention of blind assessment by Brown-Séquard that I have been able to find.

cated by a "mainstream" research scientist having undergone an assessment under conditions of ignorance. Again, intentional ignorance was selectively utilized for potentially outrageous claims.[90]

Testing Stimulants in Germanic Europe

Regarding the Brown-Séquard extract, German-speaking physiologists saw the necessity of blind assessment more urgently than did their French counterparts. In this part of Europe, the scientific concern with hypnotism and suggestion quickly became better established than in France.[91] By 1894, Auguste Forel (1848–1931), director of the famed Burghölzli Psychiatric Hospital in Zurich and well known as a brain anatomist and entomologist, made an early plea for "a concerted scientific-logical activity . . . to eliminate . . . the suggestive aspects [of a treatment] from other aspects in the healing consequences of therapy."[92] He was especially critical of the "senile-erotic ideas" of testicular extract, whose purported effects he considered to be due to "powerful suggestive factors."[93]

In 1894, Dr. Fritz Pregl at the Institute of Physiology at the University of Graz (Austria) was one of the early scientists to take up Forel's challenge and perform an experiment on the effects of testicular extract on the work performance of two students. Pregl adopted precautions to ensure that "the influence of suggestion" had been "eliminated or [was] present in equal quantities in the test persons."[94] The experiment supported the testicular extract claims as being beyond the effect of suggestion alone.

90. At least one other blind assessment on Brown-Séquard's testicular extract was performed in France. It was a single-subject cross-over design (extract, then water, then extract): see Charles Éloy, *La méthode de Brown-Séquard* (Paris: J. B. Baillière, 1893), p. 47.

91. Gauld, *History of Hypnotism* (n. 20), p. 345.

92. Auguste Forel, "Das Verhältnis gewisser therapeutischer Methoden zur Suggestion," *Zeitschrift für Hypnotismus, Suggestionstherapie, Suggestionslehre und verwandte psychologische Forschungen,* 1893/94, *2:* 385–90, quotation on p. 390. The article is based on a lecture presented in 1894 at the 60th Meeting of the German Biological Researchers and Doctors in Vienna. Forel visited Bernheim in 1887 and was considered an accomplished hypnotist. See Henri F. Ellenberger, *The Discovery of the Unconscious: The History and Evolution of Dynamic Psychiatry* (New York: Basic Books, 1970), p. 88.

93. Forel, "Verhältnis gewisser therapeutischer Methoden" (n. 92), p. 388. Forel specifically stated that "one speaks of results obtained without the patient's kowledge; but how can a patient not notice an injection? One should also make comparisons with injections of other substances" (ibid.).

94. Fritz Pregl, "Zwei weitere ergographische Versuchsreihen über die Wirkung orchitischen Extraktes," *Archiv für die gesamte Physiologie,* 1896, *62:* 379–99, quotation on p. 387. After an initial baseline assessment, one student received the extract and another

The interest in suggestion generally, and the Brown-Séquard testicular extract debate in particular, seemed to generate a more widespread movement by German physiologists toward blind assessment of various substances on healthy subjects. By 1895, a series of papers had begun to appear in German that had placebo controls in the evaluation of such substances as cola, caffeine, cocaine, mate, alcohol, and tea. For example, in 1895 one of the earliest such physiologists, Waclaw Sobierański, a Polish-born docent at the University of Marburg (Germany), explained how he preceded his experimental assays of caffeine and cocaine with saline injections.[95] He also criticized his colleagues for neglecting suggestion. He reported that he routinely used bread pills and saline injections to control for "autosuggestion," which "played a large role in healing," and that "much of the healing properties of chemical medicine is in reality attributable to autosuggestion."[96] He argued that in "order to keep psychological influence to a minimum, all subjects needed to be kept unaware of the experimental substance," kept ignorant that a deception was possible, and that all smells and appearances that could identify pills or injections needed to be "masked."[97] Placebo controls had become a method, for at least some mainstream scientists, of investigat-

received sham glycerin. Later, in a crossover manner, they received the oppoosite substances. Pregl states that the blinding precautions included mixing salt with the glycerin so that "its injection . . . more or less produced the same burning and pressure sensation" as the verum (ibid., p. 385). As a further safeguard, syringes were filled in another room that the subjects could not enter, and subjects were treated separately (ibid., p. 386).

95. Wacław Sobierański, "Über den Einfluss des pharmakologischen Mittels auf die Muskelkraft der Menschen," *Centralblatt für Physiologie,* 1896, *5:* 126–27. His research papers do not use the word *blind,* but that is clearly his intention. This particular paper studied the effects of cocaine and caffeine on muscle strength as measured by "exhaustion curves" using ergography.

96. Wacław Sobierański, "O wpływie środków farmakologicznych na siłę męśniową ludzi," *Gazeta Lekarska (Warsaw),* 2d ser., 1896, *16:* 86–95, quotation on p. 90. A summary of his research efforts was reported in a Polish-language lecture in Warsaw on 24 September 1895, though most of his scientific papers were written in German.

97. Ibid., p. 89. Even when blind assessment was acknowledged as important, its implementation was erratic. For example, in 1899 a Hanover military officer and professor named Schumburg was interested in rumors that cola kept French soldiers marching while their horses and mules were too tired to even eat. In his experiments, soldiers were given either a sham substance or cola-extract on different days; for unexplained reasons, however, when he evaluated caffeine, tea, mate, and alcohol he used no sham intervention: Wilhelm A. E. F. Schumburg, "Ueber die Bedeutung von Kola, Caffee, Thee, Mate und Alkohol für die Leistung der Muskeln," *Archiv für Anatomie und Physiologie,* 1899, *5:* 289–313, quotation on p. 293. Schumburg's other experiments include Wilhelm A. E. F. Schumburg, "Ueber die Bedeutung des Zuckers für die Leistungsfähigkeit des Menschen," *Zeitschrift für diätetische und physikalische Therapie,* 1899, *2:* 185–88. The German interest in blind assessment received a boost with the discovery of subliminal stimuli in 1908; see Gerd Gigerenzer,

ing potential "orthodox" claims for the action of substances on human subjects.

Pharmacological Testing in English-Speaking Countries

Soon after the turn of the century, the experimental psychologist and anthropologist W. H. R. Rivers (1864–1922), of Cambridge University, explicitly adopted his German colleagues' method for testing stimulants and other substances. Reflecting the prevailing Anglo-American lack of interest in the Continental preoccupation with "suggestion," Rivers stated that his concern was less suggestion and more the arousal caused by "the interest and excitement" in the experiment.[98] His experiments were "carried out with the use of control mixtures which have usually been wholly indistinguishable from those containing the active substance."[99]

Rivers's notions of blinding in pharmacological research seem to have been introduced to the United States—with only minimal effect—by such figures as H. L. Hollingsworth (1880–1956) at Columbia University, Torald Sollmann (1874–1965) of the American Medical Association's Council of Pharmacy and Chemistry, and David Macht (1882–1961) at Johns Hopkins University. All three were aware of Rivers's work and the preceding German work, though "suggestion" never seems to be mentioned in their writings. In 1912, Hollingsworth replicated Rivers's method with caffeine.[100] In 1913, a physician under Sollmann's direct supervision compared the effects of "natural" and "synthetic" sodium salicylate (two active agents) by having a group of physicians blindly administer boxes with "only a serial number."[101] The rationale for the blinding was the issue of "bias," described in terms that related to the concern with

Zeno Swijtink, Theodore Porter, Lorraine Daston, John Beatty, and Lorenz Krüger, *The Empire of Chance: How Probability Changed Science and Everyday Life* (Cambridge: Cambridge University Press, 1989), p. 87. The psychophysics movement may have also contributed to the German pharmacological interest in blind assessment. See Dehue, "Deception" (n. 60).

98. William H. R. Rivers, *The Influence of Alcohol and Other Drugs on Fatigue: The Croonian Lectures Delivered at the Royal College of Physicians in 1906* (London: Edward Arnold, 1908), p. 19.

99. Ibid., p. 20.

100. Harry Levi Hollinsworth, "The Influence of Caffein on Mental and Motor Efficiency," *Arch. Psychol.*, 1912, *22:* 1–166.

101. Albion W. Hewlett, "Clinical Effects of 'Natural' and 'Synthetic' Sodium Salicylate," *JAMA*, 1913, *61:* 319–21, quotation on p. 319. Of the 82 physicians recruited, only 27 reported back on 230 separate observations. The word *blind* does not appear in the paper. Hewlett (1874–1925) was a professor at the University of Michigan Medical School.

variances in systematic observations of astronomy.[102] In 1916 Macht, following the introspective tradition of experimental psychology, performed a series of experiments on himself and two students studying the precise quantity of analgesia produced by opium alkaloids. Again, to avoid "bias," he followed Rivers's methods and adopted "as controls normal saline and other inactive substances [which] were often substituted in place of the drug without the subject's knowledge."[103]

German Scientists Incorporate Blind Assessment into Clinical Medicine

Eventually, the German concern about suggestion moved into the arena of clinical research. The first large-scale comparative clinical trial of a conventional medical treatment involving a blind sham assessment that I have been able to uncover is a forgotten experiment from the end of the diphtheria antitoxin debate. Performed by Adolf Bingel (1879–ca. 1950), a physician at a district hospital in Braunschweig, near Hanover, the experiment sought to demonstrate that the diphtheria antitoxin, newly accepted in the profession, was in fact no better than a sham treatment.[104] Between 1911 and 1914, Bingel assigned 937 patients alternately

102. Hewlett described the rationale for the blinding as related to "the personal equations of different observers [and] the tendency to bias" (ibid., p. 319). (Also see n. 62.) In general, when the European concern for blind assessment was translated into English, the Continental preoccupation with suggestion was omitted. This can be seen in the writings of Sollmann, who was German-born and did postgraduate work in Germany. When he made an early English-language plea for blind assessment he spoke only of natural history: even "the best type of clinical reports . . . lack one important essential, namely, an adequate control of the natural course of the disease. . . . Since this cannot be controlled directly, it must be compensated indirectly. . . . The . . . method consists in the attempt to distinguish unknown preparations by their effects—the method that might be called . . . the 'blind test'" (Torald Sollmann, "The Crucial Test of Therapeutic Evidence," *JAMA*, 1917, *69:* 198–99, quotation on p. 199). Sollmann's call seems to have been mainly ignored, and it is unclear to what extent he acted on his own prescription. Harry Marks, *Progress of Experiments* (n. 3), p. 36, cites at least one other Sollmann blind assessment.

103. David I. Macht, N. B. Herman, and Charles S. Levy, "A Quantitative Study of the Analgesia Produced by Opium Alkaloids, Individually and in Combination with Each Other, in Normal Man," *J. Pharm. Exp. Ther.*, 1916, *8:* 1–37, quotation on p. 7.

104. The introduction of diphtheria antitoxin in 1894 was accompanied by much hope and also doubt. Inexact identification of the illness, difficulty with the timing of the serum's administration, and general utilization of historic or other-site comparisions when the virulence of the disease varied from season to season and place to place, made evaluation difficult and engendered lack of enthusiasm, dissension, and open skepticism. See Rothstein, *American Physicians* (n. 31); Evelynn Maxine Hammonds, "The Search for Perfect Control: A Social History of Diphtheria, 1880–1930" (Ph.D. diss., Harvard University, 1993).

to either diphtheria antitoxin serum or normal horse serum (as a sham).[105] All patients and participating physicians (except Bingel) were unaware of which treatment was genuine; the trial would probably be considered "double-blind" in today's lexicon.[106]

The German interest in suggestion and the necessity of blind evaluation for pharmacological substances reached its culmination in the work of Paul Martini (1889–1964), a Berlin medical researcher. As far as I can determine, his sixty-nine-page manual, *Methodenlehre der therapeutischen Untersuchung* (The methodology of therapeutic investigation), is medicine's first monograph to meticulously describe a methodology for controlled experimentation in clinical drug investigation. Martini's rules were designed to exclude the confounding effects of suggestion.[107] He imple-

105. The actual design of the experiment was more complex. In 1911, Bingel admitted only adults into the trial. In 1912, he lowered the age of the subjects, and by 1913 he was allocating patients regardless of age or severity of disease. Eventually, 90 percent of his patients were children. See Adolf Bingel, "Über Behandlung der Diphtherie mit gewöhnlichem Pferdeserum," *Deutsches Archiv für klinische Medizin,* 1918, *125:* 284–332.

106. Bingel's exact words were: "in order to make as objective a test as possible . . . [I asked for evaluations from] the attending physicians . . . without explaining to them the nature of the test serum" (ibid., p. 288). Bingel explicitly called his procedure a "blind" method (his quotes) and recommended its adoption by his colleagues (ibid.). He continued to perform experiments with diphtheria antitoxin even during World War II: Adolf Bingel,"Wirkt das Diphtherieheilserum bei der menschlichen Diphtheriekrankheit spezifisch durch seinen Antitoxingehalt oder unspezifisch?" *Deutsche medizinische Wochenschrift,* 1949, *74:* 101–3; idem, "Zur umstrittenen Wirkung des Di.-antitoxins beim Menschen," ibid., 1950, *47:* 1585–87. He also received explicit support for his position from another clinical trial performed on "more than 450 patients" in which "treatment was alternatively either genuine serum antitoxin or 'empty' horse serum [*Pferde-Leer-Serum*]" (A. Hottinger and D. Töpfer, "Über den Wert der Serumtherapie bei Diphtherie, insbesondere bei der malignen, toxischen Form," *Zeitschrift für Kinderheilkunde,* 1933, *54:* 505–40, quotation on p. 513). Not surprisingly, Bingel's work has been entirely forgotten, while the famous 1898 open-label diphtheria experiment conducted by Johannes A. G. Fibiger (1867–1928) at Blegdam's Hospital in Copenhagen with 488 patients is a more commemorated episode in the official histories of diphtheria and of clinical trials. (It is worth speculating that the "peculiar" results that some blind assessments produced may have contributed to the resistance to the method. Warner has pointed out a similiar predicament when orthodox medicine resisted the "numerical method . . . as a revealer of therapeutic truth" because it could present homeopathy in a favorable light [John Harley Warner, *The Therapeutic Perspective: Medical Practice, Knowledge, and Identity in America, 1820–1885* (Cambridge: Harvard University Press, 1986), pp. 202–3].)

107. His methods included rules to "exclude [*Ausschaltung,* switch off] suggestive or other irrelevant [*unsachlicher*] factors in the blinded test [*unwissentliche Versuchsanordnung*]. . . . The medications must be given to the patient in a shape or wrapping that does not permit recognition of their special character or purpose, they must be camouflaged. . . . Even during the preobservation period . . . [one must use] a fake medication treatment using inert substances" (Paul Martini, *Methodenlehre der therapeutischen Untersuchung* [Berlin: Julius Springer, 1932], p. 8).

mented his procedure in "an exaggeratedly strict fashion" in a series of experiments on drugs that treated acute symptoms (especially for angina pain) in primarily single-subject crossover studies, which would now be called "single-blind."[108]

With Martini's students carrying forward his methods, blind assessment became more frequent in 1930s Germany (though it was still far from common), with researchers using sham assessment to control for what they called "autosuggestion"[109] and the "suggestibility, the expectations of the patient and . . . the physician's personality."[110]

Blind Assessment and the Anglo-American Design of Controlled Trials

The motivation for blind assessment in Anglo-American clinical research developed mainly as an issue of experimental design and initially had little to do with the Continental concern with suggestion.[111] For Anglo-American researchers, it was technical organizational problems that were the driving force. Later, new rationales (unknowingly borrowed from the unconventional wars of an earlier era) would be deployed, but the originally structural issues were fundamental. This Anglo-American adoption of blind assessment can be described in three stages.

Concurrent Controls and Placebos

At the beginning of the twentieth century, the broad community of clinical researchers had only begun to understand the importance of concurrent controls and a systematic method of matching units of a trial

108. E.g., Paul Martini, "Klinische Untersuchung des sog. Herzhormone bei Angina pectoris," *Deutsche medizinische Wochenschrift,* 1932, *58:* 569–72, quotation on p. 570. Interestingly enough, in 1938–39 Martini applied the blind assessment method to a series of historic déjà-vu assessments of homeopathic remedies; e.g., Paul Martini, L. Bruckmer, Karl Dominicus, A. Schulte, and A. Stegemann, "Homöopathische Arzneimittel—Nachprüfungen," *Naunyn Schmeidebergs Archiv für experimentelle Pathologie,* 1939, *191:* 141–71.

109. A. Krumeich, "Klinische Prüfung der Wirkung von Arzneimitteln auf den erhöhten Blutdruck," *Deutsches Archiv für klinische Medizin,* 1933, *173:* 527–40, quotation on p. 527.

110. R. Schwenk, "Über den Wert des Histidins bei der Behandlung des Ulcus ventriculi und duodeni," *Deutsches Archiv für klinische Medizin,* 1941, *189:* 139–58, quotation on p. 139.

111. Generally speaking, American and British mainstream physicians lagged behind their Continental colleagues in conceding any ground to suggestion, hypnotism, or nonmaterial agency. An example would be the late recognition that the British Medical Association and the American Medical Association accorded hypnotism: cf. Subcommittee of the Psychological Medicine Group Committee of the British Medical Association, "Medical Use of Hypnosis," *Brit. Med. J.,* Suppl., 1955, *1:* 190–93; Council on Mental Health, "Medical Use of Hypnosis," *JAMA,* 1958, *168:* 186–89. Also see n. 101.

in living systems.[112] When researchers began to enlist patients (as opposed to soldiers and volunteers) in a no-treatment, "natural history," comparison arm of a clinical trial, they were faced with a serious problem: in the words of a famous researcher, "one cannot invite . . . [patients] for a trial, [and] obviously keep half as 'guinea-pigs,' and then hope for co-operation and good records."[113] An obvious no-treatment arm became a recruitment and retention nightmare. Patients would demand "real" treatment or seek out cointerventions. For Anglo-American clinical researchers, the initial adoption of a placebo sham in an experiment was an architectural device to create a viable and camouflaged concurrent no-treatment arm in a clinical trial. (The term *clinical trial* itself was probably coined as late as 1931.)[114] Because informed consent was still not an ethical norm, this dummy treatment could easily become a legitimate concurrent control.

The first Anglo-American experiment using placebos to provide coherence for a no-treatment group appears to have been the Michigan tuberculosis trial of 1926–31. (This trial is one of the most cited as "first blind assessment" in conventional histories of medical research.) At a tuberculosis sanatorium in Northville, near Detroit, researchers seem to have been neither aware of any earlier examples of blind assessment nor concerned with suggestion. Rather, this team felt that previous investigators could not distinguish "the notorious tendency of the disease to fluctuate naturally" from changes induced by the substance or procedure.[115] Explicitly stating that the common procedure of relying on a

112. Although there were early examples (e.g., the scurvy trial conducted by James Lind [1716–1794]), widespread attention to the issue began only with Louis Pasteur's adoption of concurrent controls in his anthrax experiments on animals (1881). Fibiger's 1898 diphtheria experiment (see n. 105) is often credited as the first carefully performed experiment on humans, where treatment was allocated impartially on entrance (by alternative assignment). But the principle did not begin to take hold for more than another generation. See Alvan R. Feinstein, *Clinical Epidemiology: The Architecture of Clinical Research* (Philadelphia: Saunders, 1985), pp. 685–86; and cf. Lilienfeld, "*Ceteris paribus*" (n. 1). The emerging perception that random assignment could eliminate bias in treatment groups (which began to have an effect in medical research with the pioneering work of Major Greenwood [1880–1949] and Udny Yule [1871–1951] on typhoid and cholera) also helped ignite an awareness of the need for equivalent control groups.

113. Austin Bradford Hill, "The Clinical Trial," *Brit. Med. Bull.*, 1951, *7* (4): 278–82, quotation on p. 281.

114. "Clinical Trials of New Remedies," *Lancet*, 1931, *2:* 304.

115. James Burns Amberson, B. T. McMahon, and Max Pinner, "A Clinical Trial of Sanocrysin in Pulmonary Tuberculosis," *Amer. Rev. Tuberc.*, 1931, *24:* 401–35, quotation on p. 429.

series of case histories was unreliable, the team adopted the emerging notion of concurrent controls. Therefore, twelve patients received sanocrysin (sodium-gold-thiosulphate, 37.4% gold) and twelve meticulously matched patients received "intravenous injections of distilled water."[116] Two of the three authors knew which group received the gold solution, as did the ward nurse, making the trial what today would be called single-blind.[117]

The second Anglo-American trial using a placebo in a no-treatment group was apparently an effort at the Cardiac Department of London Hospital, beginning in 1930. This experiment involved ninety ambulatory patients with angina pain. Seemingly unaware of Martini's concurrent work on angina, the investigators deceptively administered placebos to give the "no-treatment" comparison group in the crossover design a defined structure. A variety of sixteen drugs (including nitrates, narcotics, digitalis, and belladonna) was given, interspersed with periods of placebo administration. Like the Michigan group, these researchers were not concerned with suggestion, but rather wished to control for "the natural variations in the severity of the symptoms" that a comparison with an untreated group could detect.[118] No tested drug in this trial was better than the placebo. Several other blind assessments occurred in the 1930s that used a sham to masquerade as treatment in a no-treatment group.[119]

116. Ibid., p. 406. Treatment allotment was by "flip of the coin," and neither the word *blind* nor the word *placebo* appeared in the paper.

117. Lilienfeld's reading ("*Ceteris paribus*" [n. 1], p. 17) of the paper makes him consider this trial "double-blind," which seems plausible.

118. William Evans and Clifford Hoyle, "The Comparative Value of Drugs Used in the Continuous Treatment of Angina Pectoris," *Quart. J. Med.*, n.s., 1933, *26:* 311–38, quotation on p. 336. The report emphasized the spontaneous variation of the disease as the reason to adopt placebo control but did mention "mental suggestion" as a possible explanation for the positive results obtained through use of the placebo (ibid., p. 335). This paper used the word *placebo,* which at the time was very unusual in a research report.

119. The most important of these experiments was a series of large-scale trials concerning various treatments and preventive measures for the common cold, performed on student subjects under Harold S. Diehl (1891–1973) at the University of Minnesota. The reports included Harold S. Diehl, "Medicinal Treatment of the Common Cold," *JAMA,* 1933, *101:* 2042–49; Harold S. Diehl, A. B. Baker, and Donald W. Cowan, "Cold Vaccines: An Evaluation Based on a Controlled Study," *JAMA,* 1938, *111:* 1168–73; and Donald W. Cowan, Harold S. Diehl, and A. B. Baker, "Vitamins for the Prevention of Colds," *JAMA,* 1942, *120:* 1268–71. The placebo controls were designed to show "how much improvement should be considered as due to spontaneous recovery" (Diehl, "Medicinal Treatment" [n. 119], p. 2044). The 1933 experiment trial was "double-blind": "The ratings were made by me and independently by another physician without either of us knowing what medication had been given to the person making the report" (ibid., p. 2043). Other pre–World War II

The American Recognition of an "Active" Sham Treatment

Harry Gold (1899–1973) and a team of colleagues at Cornell University Medical School published a paper in 1937 that, especially in retrospect, introduced a shift in the American (and eventually British) understanding of the necessity for blind assessment.[120] For the first time in the English-language research literature a placebo treatment had implications beyond just disguising a no-treatment group: it could also have a quantitative medical outcome, in addition to producing information about the natural history of a disease.[121] For the first time, American researchers conceded that "ineffective" therapy could have confounding effects on expert judgment even if a physician "knew" the spontaneous course of the disease.

Gold's group tested whether the methyl xanthines (theobromine and aminophylline) were helpful for angina patients. The experiment was similiar in design and outcome to the London angina experiment mentioned earlier (although Gold claimed that it was independently conceived).[122] The interpretation had one "minor" but crucial difference.

---

examples of placebo controls used to shape a no-treatment group include Ben Z. Rappaport, Michael Zeller, and Emanuel Padnos, "Ragweed Oral Pollen Therapy Compared with Oral Placebo," *JAMA*, 1940, *115:* 25–27; and George V. LeRoy, "The Effectiveness of the Xanthine Drugs in the Treatment of Angina Pectoris," *JAMA*, 1941, *116:* 921–25 (both of these experiments took place in 1939).

120. It seems that as late as 1933, Gold and his team were still unfamiliar with and had never implemented an experiment with blind assessment. See, e.g., Harold L. Otto, Harry Gold, and Charles R. Messeloff, "Studies on Digitalis in Ambulatory Patients with Cardiac Disease," *Arch. Intern. Med.*, 1933, *52:* 725–38. In 1935, Gold co-authored a study using the method of the "blind test" (quotation marks in the original) comparing two active substances (Ella M. Hediger and Harry Gold, "U.S.P. Ether from Large Drums and Ether from Small Cans Labeled 'For Anesthesia'," *JAMA*, 1935, *104:* 2244–48, quotation on p. 2245.) "Those administering the anesthetics . . . were unaware of the source of the ether and identified the specimens in terms of code numbers." (ibid.)

121. There were already a few eloquent English statements from medical brahmins that acknowledged "suggestion" and the fallout of the Salpêtrière–Nancy debate. In his last literary work, Charcot himself conceded defeat and spoke of "a confidence, a credulity [and] receptivity of suggestion" as being responsible for the healings that occur at famous religious shrines (Jean-Martin Charcot, "The Faith-Cure," *New Rev.*, 1893, *8:* 18–31, quotation on p. 19; this appeared in French as "La foi qui guérit," *Archives de Neurologie*, 1893, *25:* 72–87). Osler repeated Charcot's argument in discussing religious healing: William Osler, "The Faith That Heals," *Brit. Med. J.*, 1910, *1:* 1470–72. But such sentiments were generally restricted to grandiose discussions of the "art" of medicine and were not seen as having research relevance.

122. Harry Gold, Nathaniel T. Kwit, and Harold Otto, "The Xanthines (Theobromine and Aminophylline) in the Treatment of Cardiac Pain," *JAMA*, 1937, *108:* 2173–79, quotation on p. 2178.

Gold mostly repeated the normal explanation of "spontaneous variations" as the cause of the equal improvement in the experimental and placebo arms; but at the end of these standard remarks, Gold's team introduced as subsidiary points an entirely new explanation for the effectiveness of sham intervention. These points included "confidence aroused in the treatment," the "encouragement afforded by any new procedure," and "a change of the medical advisor."[123] Although Gold seemed to studiously avoid the continental word *suggestion*, operationally the European justification for a placebo in blind assessment had crept into the English-language research literature.[124] Gold and his colleagues became important proponents of blind assessment after World War II,

123. Ibid., p. 2177. Actually, Gold's experiment had one other crucial difference from the London experiment. Gold was also concerned with physician bias, and his experiment could almost be considered double-blind: "in a further attempt to eliminate the possibility of bias, the questioner usually refrained from informing himself as to the agent that had been issued until after the patient's appraisal" (ibid., p. 2175). The study seems to have begun with a single-blind design and evolved into a double-blind one. It should be noted that it was Gold and his colleagues who in 1950 seem to have been the first to use the now-established phrase *double-blind test:* see Theodore Greiner, Harry Gold, McKeen Cattel, Janet Travell, Hyman Bakst, Seymour Rinzler, et al., "A Method for the Evaluation of the Effects of Drugs on Cardiac Pain in Patients with Angina on Effort," *Amer. J .Med.*, 1950, *9:* 143–55, quotation on p. 146.

124. While Gold's work eventually became critical for the acceptance of placebo controls among his peers in the biomedical community, the notion that the mind and belief could produce "medical" outcomes had already made inroads into the English-language research literature. In 1936, two single-blind clinical trials on ulcer disease used saline injections as controls. The positive results of the sham were explained as partly due to "psychic effects" (David J. Sandweiss, "Treatment of Gastroduodenal Ulcer with Histidine Monohydrochloride [Larostindin]," *JAMA*, 1936, *106:* 1452–59) or "suggestion" (C. A. Flood and C. R. Mullins, "Treatment of Peptic Ulcer by Means of Injections," *Am. J. Dig. Dis.*, 1936, *3:* 303–5). Also in 1930, a placebo was given to 40 hypertensive patients. This experiment had no active treatment. The positive outcomes were explained as "the suggestion inherent in any drug" (David Ayman, "An Evaluation of Therapeutic Results in Essential Hypertension," *JAMA*, 1930, *95* (4): 246–49, quotation on p. 249). Ayman explicitly states that this idea is a personal opinion. Additionally, the orthodox anti-quackery crusades of the 1920s often used sham controls and may have contributed to this increased mainstream recognition of the imagination and subconscious bias. For example, see discussions in: Anon., "Two Electronic Diagnoses: The Reactions of a Guinea-Pig and Sheep to the Reaction of Abrams," *JAMA*, 1922, *79* (27): 2244–48; Austin C. Lescarboura, "Our Abrams Verdict. The Electronic Reactions of Abrams and Electronic Medicine in General Found Utterly Worthless," *Scientific American*, 1924, *131* (3): 158–160, 220–22; Arthur K. Cramp, "Some Bald Facts: Professor Scholder Appeals to an Ancient Weakness," *Hygeia*, 1927, *5:* 497–99.

but Gold's impact before the war was minimal and evidence of his influence can be found in only a few experiments and discussions.[125]

The Randomized Controlled Trial

The potential of intentional ignorance to preserve the integrity of concurrent controls or to combat delusion, bias, or other psychological factors convinced only a few pre–World War II Anglo-American researchers of the necessity for treating patients "in the dark." Until well past World War II, clinical research consisted overwhelmingly of open-labeled comparative trials.[126] The impetus for the swift adoption of blind assessment into a post–World War II mantra of modern experimentation was intimately tied to the rapid introduction, acceptance, and wholesale assimilation of the fully randomized research design of R. A. Fisher (1890–1962). Blind assessment was perceived as the corollary and companion of a new methodology that could make medicine into a full-fledged "hard" science.

Fisher's insights developed at the agricultural experimental station in Rothamsted, England. They began to have a decisive impact on experimental design with his second statistical book, *The Design of Experiments* (1935), which emphasized the importance of randomization for allowing

125. The most direct pre–World War II influence of Gold seems to have been on another New York team at Mt. Sinai Hospital, which replicated Gold's experiments on the xanthines; their report's analysis followed the exact wording of the Gold paper and explicitly mentioned "confidence," "encouragement," and the patient-physician relationship: Arthur M. Master, Harry L. Jaffe, and Simon Dack, "The Drug Treatment of Angina Pectoris Due to Coronary Artery Disease," *Amer. J. Med. Sci.*, 1939, *197:* 774–82, quotation on p. 774. Another team, at Boston's Beth Israel Hospital (with connections to Harry Gold), adopted sham saline injections as a control in a rheumatoid and osteoarthritis experiment; again, their discussion saw the control as necessary for both "the tendency to natural remission in chronic arthritis . . . [and] the psychological effect of the injection itself" (Nathan Sidel and Maurice I. Abrams, "Treatment of Chronic Arthritis: Results of Vaccine Therapy with Saline Injections Used as Controls," *JAMA*, 1940, *114:* 1740–42, quotation on p. 1742. By 1942, Harold Diehl (mentioned in n. 118) had also explicitly adopted the idea that a control group could experience "psychologic effects" (Harold S. Diehl, "Abstract of Discussion," *JAMA*, 1942, *120:* 1270–71, quotation on p. 1271).

126. E.g., see Otho B. Ross, "Use of Controls in Medical Research," *JAMA*, 1951, *145:* 72–74, which presented a quantitative discussion of the lack of well-controlled comparative clinical studies in the most prestigious medical journals between January and June of 1950. Ross's criteria emphasized concurrent controls, and he introduced the idea of randomization. Although his two examples of well-controlled trials used a sham interventon for concurrent treatment, he did not explicitly mention blind assessment as a criterion for a good trial.

a measure of uncertainty and variability.[127] For Fisher, randomization allowed estimates of error and tests of significance "to be fully valid" because "the very same causes [variability and chance] that produce our real error shall also contribute the materials for computing an estimate of it."[128] Fisher's method allowed the use of probabilistic statistics, and the very indeterminateness of variability inherent in biological phenomena, to force nature to yield a mathematically precise answer. This measure of uncertainty itself allowed for causal inference.

For the medical elite struggling to make medicine more of a science, Fisher's ideas offered a method with the appearance of scientific exactitude that could imitate the determinism and objectivity of the laboratory model and mimic the mechanical and mathematical precision of hard science.[129] Clinical research could free itself from human judgment and resemble the precision of the laboratory. The perceived scientific rigor of the RCT could legitimately stake a claim to much of the terrain that had previously clung to medicine's "art." The RCT offered "the dream of the scientist who arrives at new knowledge by a completely mechanized procedure."[130]

The new methodology of assessing efficacy, standardized by statisticians and predetermined protocols, would take "a multitude of decisions out of the hands of participating investigators . . . [and] remove a series of opportunities for clinicians to frustrate" and thereby confound experiments.[131] As the historian Harry Marks has noted, for the researcher, "the improvements in experimental method offered by statisticians represented an elegant technical fix for a host of previously insoluble organizational and social problems."[132] The masked RCT could significantly

127. Much has been written on R. A. Fisher. Examples of different approaches to his influence include C. Radhakrishna Rao, "R. A. Fisher: The Founder of Modern Statistics," *Statist. Sci.*, 1992, *7:* 34–48; and F. Yates, "Sir Ronald Fisher and the Design of Experiments," *Biometrics,* 1964, *20:* 307–21.

128. R. A. Fisher, *The Design of Experiments* (Edinburgh: Oliver and Boyd, 1935), p. 47.

129. See Gerd Gigerenzer, "Probabilistic Thinking and the Fight against Subjectivity," in *The Probabilistic Revolution,* vol. 2: *Ideas in the Sciences,* ed. Gerd Gigerenzer and Mary S. Morgan (Cambridge: MIT Press, 1987), pp. 11–33.

130. Gigerenzer et al., *Empire of Chance* (n. 97), p. 211.

131. Harry M. Marks, "Notes for the Underground: The Social Organization of Therapeutic Research," in *Grand Rounds: One Hundred Years of Internal Medicine,* ed. Russell C. Maulitz and Diana E. Long (Philadelphia: University of Pennsylvania Press, 1988), pp. 319–20.

132. Ibid., p. 319. I am indebted to Dr. Marks for sharing a prepublication version of portions of *Progress of Experiment* (n. 3). Many of the insights in the RCT section of this paper are derived from this pioneering work, esp. pp. 136–63.

realign the power relationships between "art" and "science" in medicine, as it was itself a product of this transformation.[133]

Fisher's original justification for randomization was purely statistical. As mentioned earlier, randomization was a technical statistical maneuver that permitted valid null-hypothesis testing and empowered causal inference.

But stochastic statistics were insufficient justification for most physicians. Clinicians resisted random allocation. Few physicians wanted to randomly assign patients to treatment, forgo the individualization of therapy, and withhold new, promising therapies. Randomization threatened to curtail the autonomy of practitioners and the undisciplined nature of old-style clinical researchers. Austin Bradford Hill (1897–1991), architect of the 1948 British streptomycin clinical trial (generally considered the first genuinely randomized clinical trial in history), many years later confessed that he had "deliberately left out the words 'randomization' and 'random sampling numbers' at that time because . . . I might have scared them [collaborating physicians] off."[134] Hill had to counter arguments from physicians who thought something was wrong with treating patients as so many "bricks in a column," and who feared the "elimination of the responsibility of the doctor to get the individual back to health."[135] Even as late as 1955, *JAMA* could print appeals that the evaluation of drugs should be "[kept] in the hands of the general practitioner and not . . . [be] on the basis of experiments."[136]

The medical reformers who wanted to move clinical research toward the Fisherian model found insufficient support from their colleagues. Mathematical theories did not impress most physicians.[137] An additional justification specifically tailored to clinical medicine was needed. Sud-

133. Cf. David Armstrong, "Clinical Sense and Clinical Science," *Soc. Sci. Med.*, 1977, *11:* 599–601.

134. Austin Bradford Hill, "Suspended Judgement: Memories of the British Streptomycin Trial in Tuberculosis. The First Randomized Clinical Trial," *Controlled Clin. Trials*, 1990, *11:* 77–79, quotation on p. 77.

135. Austin Bradford Hill, "The Clinical Trial," *N. Engl. J. Med.*, 1952, *247* (4): 113–19, quotation on p. 118. The article cites a source for this criticism that is not traceable.

136. Robert C. Batterman, "Appraisal of New Drugs" [Letter], *JAMA*, 1955, *158:* 1547.

137. Mainland spoke of researchers' "antagonism to statistics" and experimenters as "long resistant to statistical tests" (Donald Mainland, "The Use and Misuse of Statistics in Medical Publications," *Clin. Pharmacol. Therap.*, 1960, *1:* 411–22, quotations on pp. 411, 412). Reid described physicians as feeling that "nothing [was] . . . more depressing than . . . 'the repellent symbolism' of the mathematical statistician" (D. D. Reid, "Statistics in Clinical Research," *Ann. New York Acad. Sci.*, 1950, *52:* 931–34, quotation on p. 931). Also see Donald Mainland, "The Clinical Trial—Some Difficulties and Suggestions," *J. Chron. Dis.*, 1960, *11:* 484–96.

denly, elite physicians were talking about "the fallibility of human judgement in general and of clinical . . . judgement in particular."[138] Researchers began to justify randomization in terms of needing to protect treatment arms from overzealous advocates. Previously, the taint and accusations of bias, prejudice, overenthusiasm, credulity, and delusion were reserved for deviant healers; now, what was once a fringe threat was internalized. Even the judgments of the most senior clinicians concerning the efficacy of new therapeutics were suspect. "Bias" now haunted medicine. A "placebo effect" tied to older ideas of suggestion and expectation was recognized in Anglo-American biomedicine.[139] Harry Gold's concern for the psychological impact of inert interventions suddenly became relevant.[140] Advocates of the RCT forcibly argued for expanding "hard" science into the domain of clinical research, and their new reasons were "secondhand" ones that had once pertained only to marginal medicine. Medical researchers grafted the earlier logic of intentional ignorance onto the polemic for randomization.

The new "medicalized" justification for the RCT quickly inspired both theoretical and practical reasons for also adopting blind assessment in mainstream research. On the theoretical level, the new rationale for the RCT in mainstream research was identical to the old argument for blind assessment at the fringe: randomization and blinding both shielded patients and researchers from the contamination of "knowledge."[141] Now even biomedical physicians were susceptible to bias, suggestion, and

138. Reid, "Statistics" (n. 137), p. 933. Also see, e.g., Walter Modell and Raymond W. Houde, "Factors Influencing Clinical Evaluation of Drugs," *JAMA*, 1958, *167:* 2190–99. A description of an earlier scientific situation applies here. Suddenly, the physician was no longer safe from "forces working on him that would shift his utterances out of correspondence with reality" (Steven Shapin, *A Social History of Truth: Civility and Science in Seventeenth-Century England* [Chicago: University of Chicago Press, 1996], p. xxvii).

139. This transition in understanding the "placebo" is obvious in the medical literature but is rarely discussed. See Kaptchuk, "Powerful Placebo" (n. 2).

140. Harry Gold and his colleagues were instrumental in organizing two conferences that advocated blind assessment in order to control for this newly detected placebo effect: see Conferences on Therapy, "The Use of Placebos in Therapy," *New York J. Med.*, 1946, *46:* 1718–27; Conference on Therapy, "How to Evaluate a New Drug," *Amer. J. Med.*, 1954, *17:* 722–27.

141. Hacking points out that probability has had two distinct functions: an epistemological one having to do with credibility and with "assessing reasonable degrees of belief in propositions," and a statistical one having to do with "stochastic laws of chance" (Ian Hacking, *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction, and Statistical Inference* [Cambridge: Cambridge University Press, 1975], p. 12). A similiar dichotomy could be said to apply to randomization itself, while blind assessment would have only a credibility dimension.

delusion. Intellectually, blind assessment became the perfect and necessary corollary for the RCT.

On a practical level, it was realized that the RCT benefited from blind assessment. One cannot easily make random assignments to a no-treatment group when it is clear the treatment is a dummy. Physicians and patients both tended to fudge compliance. As discussed earlier, a few pre–World War II researchers had already realized that no-treatment groups were difficult to maintain. It was also difficult just to keep the assignment and assessment personnel from becoming aware of the treatment group, unless everyone was fully blinded.[142] It was understood that "blinding is [genuinely] only possible when randomization is employed," and that randomization is only successful under blind conditions.[143] Randomization needed blind assessment. Blindness became the darkness enforcing randomization. Medical researchers were now giving new meaning, relevance, and urgency to such old refrains as "whereas I was blind, now I see."[144]

Blinding together with the RCT became the "gold standard" of science in clinical medicine. Blinding insured compliance with randomization and enhanced the elimination of bias that randomization promised for medicine. Blind assessment moved from the fringe of medicine to its very core. Suddenly, there was a "relatively new method of . . . [the] blindfold tests" that needed to be adopted by biomedicine to insure objectivity and rigor.[145] Clinical medicine needed to rely on uncontaminated evidence. By repeating enough times that blind assessment was new, most observers came to believe that intentional ignorance in re-

142. Cf. Richard Doll, "Development of Controlled Trials in Preventive and Therapeutic Medicine," *J. Biosoc. Sci.*, 1991, *23:* 265–78. One can see Hill still struggling with assuring genuine randomization in his 1951 description of the 1948 streptomycyin trial. In this trial, only the assessment radiologists were blind to intervention. In order to ensure successful randomization Hill had to resort to enforced secrecy, which can be as difficult with medical personnel as it is with anyone. Hill stated that "the allocation of the patient to treatment or control is kept secret from the clinician until after . . . [the] patient's admission. Thus he can proceed to that decision . . . without any fear of bias" (Hill, "Clinical Trial" [n. 113], p. 280).

143. Stephen Senn, "A Personal View of Some Controversies in Allocating Treatment to Patients in Clinical Trials," *Statist. Med.*, 1995, *15:* 2667. Cf. Feinstein, *Clinical Epidemiology* (n. 112), p. 688; and Peter Armitage, "The Role of Randomization in Clinical Trials," *Statist. Med.*, 1982, *1:* 347.

144. Hill, "Clinical Trial" (n. 135), p. 117.

145. Eugene F. DuBois, "The President's Address," *Trans. Assoc. Amer. Phys.*. 1939, *54:* 1–5, quotation on p. 5. DuBois was a close senior associate of Harry Gold at Cornell–New York Hospital; in his address he educated physicians regarding the availability and potential value of blind assessment, but he still was not advocating its universal adoption.

search was indeed a recent advance.[146] Medical science managed to re-create its modus operandi and its own history.

Conclusion

Modern research methodologies, including blind assessment, obviously have a complex, "context-bound" social history. The adoption of blind assessment in medicine has had as much to do with shifting political, moral, and rhetorical agendas and technical research design issues as with scientific standards of evidence. The project of using blind assessment as a tool to demarcate the boundary between material causality and mere belief has been an enterprise with far-reaching epistemological ramifications. But blind assessment has also been a vehicle to confer social authority and moral legitimacy. Intentional ignorance began as a method to challenge the "bogus" claims of unconventional medicine; some unorthodox practitioners adopted it in self-defense. At times, some nineteenth-century iconoclastic conventional medical leaders found it valuable in their polemics. Later, veiled procedures moved into psychology, psychic research, neurology, psychiatry, and pharmacology. Although the method had been available since the late eighteenth century, conventional medicine perceived its value only when other pressing considerations were at hand. At each stage of this movement, the motivation for intentional ignorance was distinct and included scientific and extra-scientific dimensions.

The history of concealed assessment has been hidden from both researchers and historians. Perhaps part of the reason for this shadowy past is the intense fervor and absolute authority with which modern biomedicine advocates it (at least when its use is possible). To use Ian Hacking's phrase, the justification is "self-authenticating."[147] Concealed history augments the appearance of an obvious transcendent truth. Questions are discouraged. "It becomes less something molded by inter-

146. The words *blind* or *double-blind* or *double unknowns* are often kept in quotation marks in biomedical journals well into the mid-1950s to denote the novelty of the technique. E.g., see Louis Lasagna, John M. von Felsinger, and Henry K. Beecher, "Drug-Induced Mood Changes in Man," *JAMA*, 1955, *157:* 1006–20; Henry K. Beecher, "Appraisal of Drugs Intended to Alter Subjective Responses, Symptoms," ibid., *158:* 399–401.

147. Ian Hacking, "Statistical Language, Statistical Truth and Statistical Reason: The Self-Authentification of a Style of Scientific Reasoning," in *The Social Dimensions of Science*, ed. Ernan McMullin (Notre Dame: University of Notre Dame Press, 1992), pp. 130–57; Hacking describes this style as follows: "the truth is what we find out in such and such a way. We recognize it as truth because of how we find it out. And how do we know that the method is good? Because it gets at the truth" (p. 135).

ests, and more an unquestioned resource upon which any interest must draw, if it ever hopes for the accolade of objectivity."[148] History disturbs the veneer of eternal validity. Perhaps an examination of blind assessment, along with other research methodologies, needs to be continued into the present. What are the scientific and extrascientific assumptions that underpin these modern methodologies?[149] Can an examination of such questions make our contemporary methods of demarcation between fact and fiction a less simple but richer process that reveals even more of the "light of truth"?

148. Ibid., p. 132.

149. Obviously, many of the authors cited in this essay—such as Marks, Gigerenzer, Porter, and Matthews—are actively engaged in this critical analysis. Many other sociologists and historians could be mentioned. On the scientific side, many debates are also raging. Concerning blind assessment, it should be mentioned that there is already a small literature concerned with critically examining the unintended consequences of blind assessment and the a priori assumptions embedded in the methodology. Important examples of this literature include Irving Kirsch and Michael J. Rosadino, "Do Double-Blind Studies with Informed Consent Yield Externally Valid Results?" *Psychopharmacology*, 1993, *110:* 437–42; Irving Kirsch and Lynne J. Weixel, "Double-Blind versus Deceptive Administration of a Placebo," *Behav. Neurosci.*, 1988, *102:* 319–23; Mauro Moscucci, Louise Byrne, Michael Weintraub, and Christopher Cox, "Blinding, Unblinding, and the Placebo Effect: An Analysis of Patients' Guesses of Treatment Assignment in a Double-Blind Clinical Trial," *Clin. Pharmacol. Therap.*, 1987, *41:* 259–65; Sydnor B. Penick and Seymour Fisher, "Drug-Set Interaction: Psychological and Physiological Effects of Epinephrine under Differential Expectations," *Psychosom. Med.*, 1965, *27:* 177–82; J. H. Noseworth, G. C. Ebers, M. K. Vandervoort, R. E. Farquhar, E. Yetisir, and R. Roberts, "The Impact of Blinding on the Results of a Randomized, Placebo-Controlled Multiple Sclerosis Clinical Trial," *Neurology*, 1994, *44:* 16–20. In fact, reading this literature was the incentive for me to undertake this examination of the history of blind assessment.