UNIVERSITY OF CALIFORNIA,
IRVINE

The Fundamental Antagonism: Science and Commerce in Medical Epistemology

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Philosophy

by

Bennett Harvey Holman

Dissertation Committee:
Professor P. Kyle Stanford, Chair
Professor Jeffrey Barrett
Associate Professor Kent Johnson
Distinguished Professor Brian Skyrms

2015

DEDICATION

To:

My father who inspired me to make the world a better place.

And

My mother who demanded that I love my work.

I learned this, at least, by my experiment: that if one advances confidently in the direction of his dreams, and endeavors to live the life which he has imagined, he will meet with a success unexpected in common hours. He will put some things behind, will pass an invisible boundary; new, universal, and more liberal laws will begin to establish themselves around and within him; or the old laws be expanded, and interpreted in his favor in a more liberal sense, and he will live with the license of a higher order of beings. In proportion as he simplifies his life, the laws of the universe will appear less complex, and solitude will not be solitude, nor poverty poverty, nor weakness weakness. If you have built castles in the air, your work need not be lost; that is where they should be. Now put the foundations under them.

Henry David Thoreau
Walden

No one who has not given the subject especial attention can imagine what impudent mendacity, what incredibly absurd claims, what impossible statements, are indulged in by those nostrum manufacturers who have made the physician their special prey… It is a false and pernicious theory that error should be left alone; that it will die out by itself. It may in time, but I am sure that error and fraud unattacked will live much longer and thrive much better and have many more followers than if exposed frequently and attacked boldly and unitedly. We need an active, persistent propaganda, a two-fold propaganda—propaganda among the laity and propaganda among physicians.

Dr. William J. Robison (1904)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the entire admittance committee at UC Irvine LPS for giving me a chance to earn a Ph. D. in a field that I had scant previous experience with; I hope this work is the first of many that pay off that gamble.

This gratitude is especially owed to P. Kyle Stanford for your continued willingness to supervise a project outside of your area of expertise. You never ceased to be able to grasp a new concept, distill its essence, and ask penetrating and challenging questions. A good deal of improvement to my ideas is indebted to your patience and willingness to first let me wander and explore unimpeded for far longer than others would deem wise and then to dissect and challenge the ultimate product of my excursions. More than one potential chapter was discarded as "less than fully baked" and as a result more perceptive ideas emerged and are included herein.

My philosophical appreciation of clarity and precision were cultivated by the entire department, but especially by the instruction and the example of Pen Maddy, Jeremy Heis, David Malament, and Brian Skyrms. Thank you for showing me that obscurity of thought and writing is not equivalent to deep insight.

My education would have been impoverished without the benefit of learning from my classmates Kevin Zollman, Rory Smead, Bill Orr, Christina Conroy, Tucker Lentz, Elliott Wagner, A.J. Packman, Bennett McNulty, Sam Fletcher, Cailin O'Connor, and Benjamin Feintzeig. This worked in particular has benefited from many long nights of conversations with my cohort members Ben Rin and Jim Weatherall. I'm privileged to count you both as friends. A special thanks is merited by Justin Bruner, my co-author for Chapter Six. Your nascent ability to label graph axes are an inspiration to all. I will miss the sixth floor computer lab and our long nights of simulations, PBR, and Golden Balls.

I would like to thank my committee members, Kyle Stanford, Kent Johnson, Jeffrey Barrett and Brian Skyrms for their feedback on earlier drafts and their high standards for inclusion. The strength of the following argument was annealed by your critiques on earlier attempts. Similar gratitude is owed to Nancy Cartwright who sat on my advancement committee and read many draft chapters. You provided invaluable guidance, feedback, expertise that was unavailable in my home department, and a relieving assurance that the dissertation was prospecting in rich philosophical territory.

From start to finish the past eight years have been made far less confusing and infinitely more productive by the administrative wisdom and genuine kindness of Patty Jones and John Sommerhauser. I will miss you both terribly.

I thank the University of Chicago Press for permission to include Chapters Five and Six of my dissertation, which were originally published in Philosophy of Science. Financial support for the completion of the dissertation was provided by the Dean's Dissertation Fellowship.

Finally, I cannot find the words to convey my appreciation for Sally Geislar. Your love, support, and willingness to think through things with me have brought me through many rough patches. I feel so amazingly fortunate to find someone who can share my enthusiasm for everything from 19th century chemistry to ornithology (even if you think cuckoos are evil). Your desire to improve our world is a constant source inspiration.

CURRICULUM VITAE

Bennett Harvey vanBenschoten Holman

2002	B.A. in Biopsychology and Cognitive Science, University of Michigan
2008	M.A. in Clinical Developmental Psychology, York University
2015	M.A. in Mathematical Behavioral Science, University of California, Irvine
2015	Ph.D. in Philosophy, University of California, Irvine

PUBLICATIONS

(Forthcoming) ——— and Bruner, J. The Problem of Intransigently Biased Agents, *Philosophy of Science.*

(Forthcoming) Why Most Sugar Pills are Not Placebos, *Philosophy of Science.*

2011 Restrictive Materialism and the Propositional Attitudes, *Philosophia*, 39, 61-70

ABSTRACT OF THE DISSERTATION

The Fundamental Antagonism: Science and Commerce in Medical Epistemology

By

Bennett Harvey vanBenschoten Holman

Doctor of Philosophy in Philosophy

University of California, Irvine, 2015

Professor P. Kyle Stanford, Chair

I consider the claims made by medical ethicists that funding by pharmaceutical companies threaten the integrity of medical research and the claims of philosophers of science that evidence-based medicine can provide a sound epistemic foundation on which to base medical treatment decisions. Drawing on both game theory and medical history, I argue that both medical ethicists and philosophers of science have missed crucial aspects of medical research. I show that both veritistic and commercial aims are enduring and entrenched aspects of medical research. Because these two drives are perpetually pulling medical research in different directions, I identify the resultant conflict as the fundamental antagonism

The primary task of the dissertation is to provide a framework that incorporates both drivers of medical research. Specifically, I argue that medical research is best conceived of as an asymmetric arms race. Such a dynamic is typified by a series of moves and countermoves between competing parties who are adjusting to one another's behavior, in this case between those who seek to make medical practice more responsive to good evidence and those whose primary motivations are instead commercial in character.

Such a model presents three challenges to standard evidential hierarchies which equate epistemic reliability with methodological rigor. The first is to show that reliability and rigor can (and do) come apart as a result of the countermeasures employed by manufactures. This fact suggests that in considering policy proposals to improve epistemic reliability, it is robustness (i.e. resistance to manipulation) that should be the crucial desideratum. The second consequence is a reorientation of medical epistemology. One of the primary strategies that manufacturers have employed is to manipulate the dissemination of information. A focus on an isolated knower obscures the impact that industry has in shaping what information is available. To address these problems medical knowledge must be understood from a social epistemological framework. Finally, and most importantly, the arms race account suggests that the goal of identifying the perfect experimental design or inference pattern is chimerical. There is no final resolution to the fundamental antagonism between commercial and scientific forces. There is only a next move.

Preface

In October of 2010, my father underwent a dual lung-transplant. When he regained consciousness, communication was difficult. His air passages were filled with the tube that was mechanically breathing for him. His arms were weak and plugged with more tubes bringing in antibiotics, immunosuppressants, and pain killers. His lack of fine motor control made writing nearly impossible and so we were often left playing games of charades. It took my family two progressively more frustrating hours to figure out that my father's first concern, after living through a heroic 17-hour operation, was to make sure the nurses had not stolen his airline tickets to Cuba while he was unconscious. At first we thought it was my dad's odd sense of humor, but we would soon find out that it was part of a psychotic state. I had feared my father would lose his life during the operation; it had never occurred to me that he would lose his mind.

At times he appeared to be completely in touch with reality and when he wasn't, most of his hallucinations were mundane and merely annoying. His most frequent delusion centered around the conviction that he was in the hallway of the hospital. As a result he was generally frustrated that we, his beloved family, weren't demanding that he receive his own room. Some of the hallucinations were paranoid, such as when he asked my brother to kill me because he thought that I was holding him prisoner. Sadly, because the paranoid delusions seemed to gradually predominate, he got very little sleep his first night. When the doctors arrived to do their rounds in the morning they classified his collection of symptoms as intensive care unit (ICU) psychosis. I was relieved to find out that the condition would self-remit in about a week. On the other hand, until he regained lucidity, the doctors wanted to put him on Seroquel, an atypical antipsychotic.

I had been delegated his medical power of attorney and, given his mental state, I was now in charge of making his choices for him. Most of the drugs my father was on were far beyond my ken, but I had a degree in biopsychology and clinical psychology. I had studied the pharmacological actions of

antipsychotics and I knew how powerful of a drug Seroquel is. I didn't understand why they wanted to give it to my dad when he was in such a fragile state, especially since the drug is approved for a disorder he didn't have.

When his doctors came around in the afternoon, I had a chance to speak with them about their treatment rationale. Empirical work has shown that the longer a person is delusional, the worse their prognosis is. The doctors felt that the psychotic break that characterizes ICU psychosis was similar to a self-remitting brief schizophrenic episode, and in their clinical experience antipsychotics shortened the duration of the psychosis. In addition, they worried about the unpredictable nature of patient behavior during psychotic episodes; for example, in a bout of confusion, my dad could try to remove the intubation that was keeping him breathing. Finally, his paranoia was preventing him from sleeping and rest was crucial at this precarious stage.

The doctors' concerns were reasonable, but I was still hesitant. I asked if there was any evidence that supported their recommended course of treatment. They assured me that it was standard procedure, but after further obstinacy on my part, they begrudgingly agreed to send over what studies they could find and return at the end of the evening to finalize a course of action if my dad's delusions did not abate and/or he remained unable to sleep. Shortly thereafter, a doctor brought me three articles, none of which actually addressed the question of whether antipsychotics improved patient outcomes.

From where I stood, I saw no evidence to support the use of an antipsychotic. There were no relevant clinical trials and the doctors' clinical impressions struck me as extremely unreliable. Quickly self-remitting diseases are prone to producing clinical lore that is unrelated to actual clinical efficacy. The correlation between longer duration of the psychotic episode and poorer outcomes was worrisome, but poorer initial neurological functioning may have precipitated both susceptibility to ICU psychosis and

worse outcomes. Further, there was no experimental evidence that atypical antipsychotics reduced the duration of the psychotic episode.

In contrast, the doctors' concerns about disorientation were very real; my dad had already tried to extubate himself that evening. I also recognized his need for rest, since he hadn't had sustained sleep since he had regained consciousness. The threat of extubation could be easily dealt with: he was weak from the surgery and so it was easy to physically restrain him from pulling out his tube. I was at his bedside 14 to 16 hours a day and on top of that he had a dedicated nurse that sat outside the door of his room. As for sleep aids, benzodiazepines (which have both anti-anxiety and sedative effects) were contraindicated because of the other treatments he was receiving, but as far as I could tell trazodone (an old antidepressant with anti-anxiety and sedative effects) was not.

It was not without trepidation that I told his doctors that I would approve of the use of trazodone, but not Seroquel for a sleep aid. There are constant reminders of the fragility of life in an ICU. We had listened to the disturbing sounds of the woman in the next room as she died that morning and the waiting room was frequently filled with grieving family members. When the time for a decision came, I simply saw no evidence that there was a benefit of using Seroquel over and above its sedative properties, and I perceived it to have greater risks than trazodone. This was partly because trazodone seemed like a "weaker" drug and partly because it had been in use for 30 years at that time, as opposed to the 10 years during which Seroquel had been in circulation. I hoped that this meant there would be fewer unknown interactions. In addition, Seroquel was still under patent and being heavily promoted, and I was unsettled by the fact that every doctor I spoke with recommended Seroquel by name. In one discussion a doctor reassured me that unlike first-generation antipsychotics, second-generation antipsychotics were much milder and did not carry with them any risks for short-term usage. By underplaying Seroquel's risk profile, the doctors lost my confidence.

Of course, I wasn't terribly confident in my decision and there were certainly trials that could have been designed to answer the questions I had, but choices like this do not wait patiently on the shelf. It was gut wrenching to go against the doctors' recommendation with my dad's life in the balance, but given the constant introduction of new treatments, the cutting edge of medicine will always be fraught with life and death choices that are made with incomplete information, and often on the basis of clinical experience.

As the days passed, there was a reduction in psychotic episodes and—most importantly—an increase in full nights of sleep, mostly unaided by the trazodone. I felt vindicated as I watched my dad retreat from the line between life and death. However, I was reminded of the frailty of such experiences when a doctor returned after being out of the ward for several days. The doctor had shown sincere concern for my father's health, and after the initial diagnosis of ICU psychosis he had graciously spent a significant amount of time explaining to me why the doctors felt that Seroquel was indicated. When he came back to the ICU he greeted me and looked at my dad's charts. He commented that I must now be glad I had okayed the Seroquel. Upon discovering that my dad's condition was primarily the result of not intervening, he closed his eyes, somberly shook his head, and gravely informed me how fortunate I was that my dad was still alive. The outcome he judged as having confirmed the success of Seroquel did not apply to the result of non-intervention.

Though my dad's delusions ceased completely, he was still in the ICU two weeks later when I overheard two nurses discussing another patient. The first was telling the second the signs of ICU psychosis, that it was an underdiagnosed condition, and that it was important to recognize and treat for a number of reasons. The first reason the nurse cited was an association between duration of psychosis and prognosis. The second was that the hospital was being considered as a possible site for a trial to study ICU psychosis, and that its selection for this trial would in part hinge on the hospital

demonstrating that it had a sufficient number of such cases. I had just discovered firsthand the need for such information and this study might well have been an important attempt to get answers to the questions I had blindly wrestled with. On the other hand, pharmaceutical companies have been known to conduct "clinical trials" in order to get doctors used to writing prescriptions they wouldn't otherwise write.

I was never able to find out which type of study was being considered while my dad was in the hospital, but the latter wouldn't have been atypical for AstraZeneca (the manufacturer of Seroquel). One year after my dad was hospitalized, it was alleged that AstraZeneca had:

> through the use of its own sales force and a group of highly compensated outside physicians, promoted Seroquel to elderly patients, children and prisoners. Many of these patients suffer from a wide spectrum of psychiatric mood and other disorders not normally considered by the FDA [Food and Drug Administration] to be appropriate for treatment with Seroquel. (Wetta vs. AstraZeneca, 12)

To be clear, the FDA approves drugs for certain indications, but it does not regulate medical practice. Once approved, doctors can use drugs however they choose. The illegality lay in the *promotion* of drugs for a purpose that has not been demonstrated to be safe and effective. Despite the fact that AstraZeneca was prohibited from promoting its drugs for off-label usage, the company donned the guise of science and "conducted what is known in the industry as 'investigator initiated trials' (IIT's). IIT's funded by AstraZeneca have been used to promote off-label use of Seroquel for… sleep deprivation and anxiety" (Wetta vs. AstraZeneca, 13-14).

As I read through the complaint against AstraZeneca, I wondered whether my father's doctors had been to the commercially funded "education seminars" that had promoted Seroquel for off-label uses. I also wondered whether they had been so insistent to prescribe my dad Seroquel as part of a push to land their clinical trial. Whatever their motivations, 45 days after his surgery, my dad left the hospital alive. That fact alone doesn't indicate whether this outcome was because of my decision or in

spite of it, but whatever the case may be, one thing the complaint made clear is that there are significant risks involved with short-term uses that my dad's doctors were unaware of, including the fact that Seroquel "predisposes non-ambulatory patients to pneumonia and other respiratory ailments" (Wetta vs. AstraZeneca, 13). Not a very attractive side-effect profile for a bed-ridden dual lung-transplant recipient.

We may charitably assume that my dad's doctors thought Seroquel was the right course of treatment given his situation, but let's also consider for the moment that they were wrong. Had they written the prescription then, two things would have happened: first, my father's chance of recovery would have worsened; second, AstraZeneca's sales total would have increased. It is this kind of conflict that, I will argue, leads to the fundamental antagonism in medical research, and that this dissertation will aim to understand. Though the discussion may at times be abstruse, it should not be lost on the reader that the state of medical knowledge is never far removed from the difference between having a father and having a funeral.

CHAPTER 1

The End is Nigh

A Beginner's Guide to Pharmageddon and Medical Eschatology

1.0 Pharmageddon

> *But actually, he thought as he readjusted the Ministry of Plenty's figures, it was not even forgery. It was merely the substitution of one piece of nonsense for another. Most of the material you were dealing with had no connection with anything in the real world, not even the kind of connection that was contained in a direct lie. Statistics were just as much a fantasy in their original version as in their rectified version… Very likely no boots had been produced at all. Likelier still, nobody knew how many boots had been produced, much less cared. All one knew was that every quarter an astronomical numbers of boots were produced on paper, while half of Oceana went barefoot. And so it was with every class of recorded fact, great or small. Everything faded away into a shadow world in which, finally, even the date of the year had become uncertain.*
>
> (Orwell, 1949, p. 41–42)

The most Orwellian version of what happened with my dad was that that despite the fact that Seroquel threatened his life, AstraZeneca promoted it for ICU psychosis on the basis of poorly run studies that it funded and then disseminated at educational events it paid for. My dad's doctors were highly educated dupes, foolishly trusting the information fed to them. In reading the more apocalyptic accounts of our current situation in medical epistemology, it is hard not to feel this kind of despair. Historian of psychopharmacology David Healy recently referred to the state of current medical practice as "pharmageddon" because it has gotten to the point where "science and progress have become marketing terms and where benefits accrue to companies even while patients suffer harm" (Healy, 2012, p. 11).

By Healy's lights pharmaceutical companies dominate the medical sphere. They create diseases that do not exist and put out inferior products to replace cheap and effective remedies. To promote their wares, pharmaceutical companies have leading academics assume the nominal authorship of articles they did not write about studies they did not conduct. These ghostwritten, company-designed, -run, and -analyzed trials are then published in top medical journals. Companies pay these leading lights of medicine to "educate" the rank and file doctor about this "new medical breakthrough." They crush the few remaining dissenters by threatening to stop funding those in their employ, ruining the careers to those who are not, and threatening legal action against anyone that dares to broadcast a heretical

message. We are in a situation where, every quarter, an astronomical number of breakthrough drugs are created on paper, while millions of patients become steadily less healthy. Ultimately, I will argue that this is a jaundiced view, but such concerns are not unfounded or insignificant and any view that ignored them would be as incomplete as one that focuses only on such problems. In the following chapters I will begin to integrate such issues into a comprehensive understanding of medical epistemology, but this chapter will be primarily devoted to identifying various ways in which commercial practices threaten the veritistic aims of medical research.

1.1 A depressing case study

As a picture of pharmageddon, consider Forest Laboratory's antidepressant Celexa (citalopram).[1] At the time that Forest submitted an application for Celaxa's pediatric use, two studies had been completed, one in 1996 and a second in 2002. The second study (Wagner et al., 2004) supported the use of Celexa; however, though it appeared to have been authored by independent academics, it was actually ghostwritten by Forest employees and quickly published. The earlier study, though complete, was not published because the data showed the drug to be ineffective and to increase the risk of suicide among children. As a result, the FDA quietly rejected Forest's application for Celexa's pediatric use. The negative study was only disseminated to a small group of top executives, thus creating the false impression that the only evidence was positive. The study was not disseminated to Forest's sales department and professional affairs department, leading them to repeatedly claim that all available evidence supported the use of Celexa for pediatric depression.

While Forest was prohibited from advertising Celexa for pediatric use, the FDA does not regulate what doctors say about their own research. Thus, instead of promoting the product *per se*,

[1] The following abuses are detailed in the lawsuit filed against Forest by whistleblower Christopher Gobble filed Feb 8, 2010. Forest pled guilty and settled the lawsuit, paying $313,000,000 in damages.

Forest actively promoted Wagner et al.'s (2004) study at conferences, continuing medical education seminars, and other venues. The company paid nearly 20,000 doctors to serve on a board as consultants. Internal documents definitively show that the purpose of this was to increase prescriptions to children. In fact, the company conducted analyses in order to assure themselves that the money they were paying doctors on the board resulted in a satisfactory "return on investment," though Forest eventually discontinued official analysis to avoid further memorandizing the true (sales) purpose of the board. Additionally, Forest conducted an "efficacy study" whose true purpose was to get over 2,000 participating doctors accustomed to writing prescriptions for their products. Finally, Forest arranged for their sales reps to go on "preceptorships" (i.e. to follow around doctors and "learn" about clinical practice). In reality, the reps tried to influence prescription practices by suggesting drug treatments using company products; they were required to fill out return-on-investment forms before paying doctors. Finally, after the existence of the negative study became an open secret, Forest published it (Von Knorring et al., 2006) to avoid further charges of withholding negative results.

The example of Forest is highlighted here not because it is uniquely bad, but because it is altogether too common. Similar problems arose with GlaxoSmithKline's product Paxil (paroxetine). In this case, leaked internal memos discussed their own negative study of pediatric use, finding it imperative to "effectively manage the dissemination of these data in order to minimize any potential negative impact" (Kondoro and Sibbald, 2004, p. 783). This was done by withholding from publication study number 377 and publicly claiming that study 329 showed that "paroxetine is generally well tolerated and effective for major depression in adolescents" (Keller et al., 2001, p. 762). The internal memo showed that the company knew that this was unsupported by the data, but found it "commercially unacceptable to include a statement that efficacy had not been demonstrated, as this would undermine the commercial profile of paroxetine" (Kondoro and Sibbald, 2004, p. 783). In other

words, if these findings were published and understood, doctors would stop prescribing children drugs that did not work.[2]

Eli Lilly, maker of Prozac (fluoxetine), dealt with its suicide problem by coding suicides that happened after recruitment for its study, but *before the beginning of the trial,* as if the suicides had occurred in the placebo group (Healy, 2006). The same coding paradigm was adopted by GlaxoSmithKline for Paxil and by Pfizer for Zoloft. In each case, this allowed the company to state that while there were more suicidal acts by participates who were on the drug, the difference between suicidal acts in the treatment group and the control group was not statistically significant and thus could not be attributed to the drug (Healy, 2012). Having been an expert witness in a lawsuit against the company, Healy had access to unpublished data and an inside look at the internal working of pharmaceutical companies.

Beginning in 1999, Healy bucked conventional wisdom and made it a personal cause to publicize the dangers of antidepressants. Friends called him and said that they had been told they should no longer be associated with him (Healy, 2004, p. 165); he was publically ridiculed (for example, Pfizer called his work "junk science"), and he had a major job offer rescinded by the University of Toronto after speaking on the potential relationship between selective serotonin re-uptake inhibitors (SSRI) antidepressants and suicide in 2001. Toronto denied that Lilly being the largest contributor to the department affected their decision. Though Healy was virtually alone for years, he has since been vindicated. The British advisory agency eventually issued guidelines against using antidepressants other than Prozac for children, and in 2004 the FDA added a black box warning that SSRIs may cause suicide in children.

Nor should it be thought that such problems are confined to SSRIs or to psychiatry. When asked to comment on the significance of the case against Seroxat (the British name for Paxil), Dr. Shooter,

[2] For an extended account of the case against GlaxoSmithKline see Bass (2008).

president of the Royal College of Psychiatrists, noted that "It has huge implications. I think once again we're seeing the SSRIs being the focus for something much wider in psychiatry and we're seeing psychiatry being the focus for something much much wider in medicine as whole… [this has ramifications] right through medicine" (Shooter, 2004). Indeed, a sample list of other products promoted by some of the same tactics that got Forest in trouble include: Merck's arthritis drug Vioxx; AstraZeneca's antipsychotic Seroquel; Cephalon's anticonvulsant Gabitril, narcolepsy treatment Provigil, and pain medication Actiq; Johnson & Johnson's migraine treatment Topamax; and Roche's antiviral Tamiflu.

If one is disposed to think of medical research as a bastion of the good and pure, where scientists and doctors labor over their benches in hopes of discovering the cure for a scourge of humanity, then reality provides a rude awakening. The horrors of pharmageddon include: disease-mongering; a low bar for obtaining a patent/FDA approval; an imbalanced confluence of interests; the manipulation of doctors' medical judgment; pharmaceutical companies assuming the form, but not the character, of scientific rigor; and the emergence of an industry to support them. As we will see in the next chapter, philosophical accounts that do not take these pressures into account are apt to draw the wrong lessons from medical history. Below, I consider each in its own right.

1.2 Disease-mongering: The threat to the reliability of diagnostic categories

In the 1970s, osteoporosis was a rare disorder that involved the dramatic thinning of bone structures and a significantly increased risk of broken bones. With the advent of dual-energy X-ray absorption scans came the ability to quantify bone density. Osteoporosis became defined based on a bone density of less than 2.5 standard deviations below average, and a new disease (osteopenia) was defined for women as 1 standard deviation below average. The average was established by looking at bone density of women in their 20s. Given that bone density naturally decreases over time, a woman in her 60s

stands to have less dense bones than the average 20 year old; only with the new definitions did this decreased density become an illness. Indeed, the redefinition meant that one third of post-menopausal women are afflicted with osteoporosis (Bassett, 2000, as cited in Healy, 2012) and half of all women will meet the criteria for osteopenia by age 52 (Abramson, 2004). In contrast to sources connected to drug companies, independent sources note that bone density is only one of several factors that predict fractures and that increasing muscle strength and balance and ceasing smoking are equally, if not more, important interventions.

While it is one thing to alert women to possible health risks, it is quite another to pathologize them. Osteoporosis is merely one example of diseases that have been created out of thin air or had their boundaries radically expanded because of marketing considerations; restless leg syndrome, over-active bladder, gastroesophageal reflux disease, irritable bowel syndrome, and bipolar disorder have all gone from the marketing department into our daily lexicon (Elliot, 2010).

1.3 Me-too drugs: A threat to the scrutability of medical knowledge

For a person steeped in the rhetoric of medical breakthroughs, one of the most surprising findings is how few drugs are actually significant improvements. From 2000 to 2004, 282 of the 314 drugs (roughly 90%) approved by the FDA were not classified as innovative. While roughly 10% of the non-innovative category were old drugs that had been modified to yield significant improvement, the vast majority were "me-too" drugs (Angell, 2004; cf. Goozner, 2004, esp. Chapter 8; Healy, 2012, p. 33); that is, chemicals rush-produced by competitors that are different enough from blockbuster drugs to win patents, but similar enough to achieve the same therapeutic effects.

Indeed, when faced with expiring patents, companies sometimes create their own me-too drugs and attempt to switch patients from the old drugs to the new drugs to keep profits high, as when AstraZeneca patented an isomer of Prilosec and proceeded to market Nexium. Chemicals come in two

forms, which are mirror images of each other (enantiomers). Prilosec contains both of the enantiomers of the omeprazole molecule, while Nexium only contains one. Even though Nexium is literally half of Prilosec, it was awarded a patent and FDA approval as a new drug. Sometimes companies are able to eliminate drug side effects by removing the offending enantiomer and selling the other, as when Aventis eliminated the side effects of its allergy drug Seldane and created Allegra. This was not the case for Nexium, a product idea created by the AstraZeneca marketing department's "Operation Sharkfin." Nexium was developed to maintain profit margins; heavy marketing to doctors resulted in patients being shifted over to Nexium before the patent on Prilosec expired and prices dropped (Harris, 2002). Forest Laboratories engaged in the same strategy in attempting to move patients from Celexa to its isomer Lexapro.

Similar ploys include patenting the drug's active metabolites or creating a new brand and use for an unpatented chemical. For example, fluoxetine costs $8.02/pill when it is given as Sarafem for premenstrual dysphoric disorder, and $0.52/pill when it is given as Prozac for depression.[3] Note that both of these chemically indistinguishable pills are made by Eli Lilly, so it is not a matter of generic fluoxetine ($0.15/ pill) vs. name-brand Prozac; the only difference here is that Sarafem's pill casing is pink and purple while Prozac's is green and white. Marketing specialist Vince Parry explained the justification for the creation of Sarafem as follows: "By changing the brand name from Prozac to Sarafem—packaged in a lavender and pink—Lilly created a brand that better aligned with the personality of the condition for a hand-in-glove fit" (as quoted in Moynihan & Cassels, 2005, p. 114). That is, Eli Lilly found a way to charge women a $7.50 premium by creating a new product from an old chemical.

Because there can be absolutely no therapeutic advantage to Sarafem over Prozac, it is a particularly good example of commercial influences in medicine. That said, the effects of me-too drugs

[3] Price information from Pharmacychecker.com on 2/6/13

go well beyond economic considerations. From an epistemic standpoint, the surfeit of new products, where all are promoted with equal enthusiasm, makes it far more difficult for a doctor to personally stay up to date or to separate true therapeutic breakthroughs from drugs that are merely promoted as such.

1.4 Advertising, reprints, and lawsuits: A threat to the reliability of medical journals

A frequent complaint is that journals, or more specifically journal editors, are serving the interests of pharmaceutical companies in two ways. The first is by uncritically publishing industry studies. Consonant with this is the perception of pharmaceutical companies that lower-rung journals will accept anything for publication (Rost, 2006, p. 188). Sometimes, the conflicts are insidious; for instance, when a paid consultant of Forest Laboratories published a study he authored, in a journal he edited, claiming that Forest's new product Lexapro was superior to its isomer Celexa (Kassirer, 2005, p. 32). However, the high-end journals have been called in to question as well.

The ghostwritten Paxil study 329 was published in the top journal for pediatric psychiatry despite making claims in the abstract that were unsupported by the data (Bass, 2008). Similarly, the prestigious *New England Journal of Medicine (NEJM)* published a highly laudatory review article of COX-2 inhibitors (specifically Celebrex and Vioxx) when the drugs began receiving negative press. This particular study highlights a number of the complexities inherent in medical publishing, so it is worth considering in greater detail (cf. Abramson, 2004, pp. 23–38).

Vioxx and Celebrex were marketed as super-aspirins that would prevent inflammation around joints without causing the gastrointestinal problems associated with traditional of non-steroidal anti-inflammatory drugs ((NSAIDs) e.g., ibuprofen). This image was created in numerous publications by industry-funded researchers who inflated the risks of traditional NSAIDs and implied that a new generation of drugs would be a boon to consumer safety. Despite the commercial hype, Vioxx and Celebrex were neither safer nor more effective than traditional pain-relievers. Accordingly, the FDA

prohibited their manufacturers from making comparative claims as these conclusions were unsupported by the data. Despite their being neither better nor safer than NSAIDs, however, COX-2 inhibitors had one of the most profitable roll-outs in pharmaceutical history, generating billions in sales (Goozner, 2004, pp. 224–229).

The marketing for these products constituted such a continued and blatant violation of FDA guidelines that the FDA took the drastic step of mandating that Pharmacia, the manufacturer of Celebrex, send a "dear doctor" letter to every physician in the country stating that Celebrex's advertising had violated the FDA's mandate and made unsubstantiated comparative claims regarding the reduced rate of serious gastrointestinal toxicity (e.g. bleeding, ulceration, or perforation of the stomach). The reason for the discrepancy between the public perception can be traced to how Pharmacia handled the publication of their trial results.

In the case of Celebrex, the CLASS study ran for a year and was intended to demonstrate a lower rate of ulcers. It failed to provide this outcome; however, instead of reporting the trial as designed, Pharmacia reported the results of carefully chosen post-hoc analyses that yielded the desired conclusion. To obtain the intended results, the company reported only from the first six months and published the study in *JAMA* as if it were the full experiment. Further, even in the six months' worth of data there was no significant difference in serious gastrointestinal complications between patients taking Celebrex and those on other NSAIDS. To obtain a result that favored Celebrex, Pharmacia created a new variable that included serious gastrointestinal complications, and then performed the comparative analysis. Again, this was reported as if it were the original design of the experiment. It was on the basis of these results, and the prestige afforded by publishing in *JAMA,* that Pharmacia was able to hype its new drug.

The VIGOR study of Vioxx, published in the *NEJM,* was also subtly manipulated, but by design. Specifically, over half of the patients enrolled in the study were taking steroids, though such a

combination of steroids and COX-2 inhibitors is rare in the actual patient population. Outside of this subpopulation there was no decrease in gastrointestinal complications, but by filling the patient pool with a disproportionate number of patients taking steroids and then failing to distinguish between them and other patients in the analyses, the authors were able to claim that the decrease identified was applicable to the general population. The authors also downplayed the increase in serious cardiovascular complications caused by Vioxx, merely recommending that people with a history of heart problems should also take a prophylactic dose of aspirin. In contrast, the FDA noted that the increase in cardiovascular problems occurred even in patients with no history of heart trouble, and concluded that NSAIDs should be favored to Vioxx.

While the FDA can regulate official statements made by pharmaceutical companies, academic researchers are not regulated even if they receive industry funding. For example, four months after Pharmacia sent their "dear doctor" letter, the *NEJM* violated its own policy and published a "review article" (i.e., an article summarizing the available data) written by two authors who had received grant support and served as consultants to Pharmacia and Merck (the manufacturer of Vioxx). Though Pharmacia was legally prohibited from claiming any safety benefit, the FDA could not stop the authors from stating that Celebrex causes "significantly fewer serious gastrointestinal events than does treatment with non-selective NSAIDs" (Fitzgerald & Patrono, 2001, p. 440). Similarly, though Merck had been warned by the FDA not to minimize the increased risk of cardiovascular events caused by Vioxx (e.g., strokes, heart attacks, etc.), the FDA could not stop the authors from presenting such a finding as a result that "may represent the play of chance." Such a conclusion is the exact opposite of that supported by the data. The difference was statistically significant for both cardiovascular events generally ($p<.05$) and heart attacks in particular ($p<.01$). Vioxx would ultimately be pulled from the market in 2004 for inducing serious cardiovascular events, but in the meantime the *NEJM* helped stem the flow of bad press and sell a lot more product.

It is unclear why editors allowed such conclusions to be published, especially given the fact that the authors had stated conflicts of interests. Moreover, a bit more diligence would have led the editors to the FDA website, where the full results of the studies were available. The discrepancies between the published and unpublished data should have been salient given the recent "dear doctor" letter. Such failures of due diligence lead to the concern that even the top medical journals "have devolved into information laundering operations for the pharmaceutical industry" (Horton, 2004, p. 8).[4]

The lax treatment given to favorable articles is put into stark relief by the reluctance or refusal to publish critical pieces (Kassirer, 2005, pp. 89–91). For example, when researchers tried to publish a paper on the selective reporting of Paxil trial 329, *The Lancet* demurred after sending the article to GlaxoSmithKline for comment. A highly edited version of the paper was published by *The British Medical Journal,* but the publication came years later and after several rounds of edits and input from the journal's legal department (Healy, 2012, pp. 123–124). Marcia Angell (former editor in chief of *NEJM*) has explained that pharmaceutical companies do not even have to explicitly be involved; simply the threat of legal action and potential loss of advertising revenue can lead journal editors to self-censor (Abramson, 2004, p. 113).[5]

The flip side is that major trials with positive findings can bring in upwards of a million dollars in revenue from reprints bought by the pharmaceutical company and distributed to doctors.[6] Though the *NEJM* later disclaimed the results of the study used to promote Vioxx, they kept the roughly $750,000

[4] Horton has been editor in chief of *The Lancet* since 1992. Further, Horton is not alone in his concern; similar views have been expressed by Richard Smith (2005), editor of the *British Medical Journal*, and Marcia Angell (2004) and Jerome Kassierer (2005), both whom served as editor in chief of *NEJM*.

[5] The loss of advertising is a significant possibility. In 1992, when the *Annals of Internal Medicine* ran an article that demonstrated the false claims made in contemporary journal ads, the journal lost $1.5 million from cancelled ad contracts (Altman, 1999).

[6] Negative articles presumably bring in nothing close to this revenue. As a personal anecdote, when I sat down with Robert Rosenthal to discuss his results on nonpublication of antidepressant research, he handed me one of the free print copies he had received when the article was published nearly two years prior. Even though it was one of the most widely cited papers on the topic, one of the lead authors was still going through his free copies. It would be very beneficial to have hard data on the actual costs and benefits of reprints for positive and negative findings.

Merck paid them for reprints (Elliott, 2010). Even when the truth comes out eventually, the time discrepancy between publishing laudatory and critical pieces leads to a commercially favorable distortion in available knowledge. At best, this translates into a misuse of financial assets and unnecessary suffering. In the case of Vioxx, from its 1999 release it has been estimated to have caused 140,000 cases of serious heart problems and 38,000 deaths, a good portion of which occurred after the *NEJM* 2001 review (Carlat, 2010; Elliott, 2010, p. 104).[7]

1.5 Gifts, detailers, and key opinion leaders: A threat to the reliability of continuing medical education

Doctors, and American doctors in particular, are some of the world's largest consumers. They neither ingest the product they consume nor think of themselves this way, but via their prescriptions they control the spending of billions of dollars. Thus, it should come as no surprise that this elite group of mega-consumers is heavily marketed to. They receive countless advertisements through the mail, and journals they have not subscribed to. They walk away from conferences with bags "weighed down with disposable cameras, textbooks, long-distance phone cards, mugs, clocks, and innumerable pens—all free of charge" (Carlat, 2010, p. 104). Every company employs an army of detailers that travel between doctors' offices, ostensibly to keep these doctors up to date on the latest drugs (Brody, 2007; Elliott, 2010). This education invariably means that prescriptions are written for the latest products offered by these companies, and results in less effective and more expensive treatments (Griffith, 1999; Lexchin, 1989; Moynihan, 2003; Wazana, 2000). The industry has even invented a typology of doctors so that detailers can individualize sales pitches (Healy, 2012, p. 55).

More recently, companies have begun to employ doctors to give presentations to each other and give feedback to the company as consultants. As noted above, Forest employed tens of thousands

[7] It should be noted that journal editors, especially of elite journals, have been at the forefront of several calls for reform.

of doctors in their promotion of Celexa. Merck found that even though they paid more money to have doctors give a presentation, a doctor that sat through a presentation from a peer wrote an average of $623.55 worth of additional prescriptions for Vioxx, whereas doctors listening to drug reps only wrote an additional $165.87 (Carlat, 2010, p. 126). Such tracking is made possible because pharmacies sell doctors' prescription data to pharmaceutical companies. The right for pharmacies to do so was outlawed in some states, but the Supreme Court ruled such laws unconstitutional.

Merck is not alone in paying doctors to give presentations; pharmaceutical companies spend one-third of their advertising budget on paying doctors to speak (Elliott, 2010, p. 78). In 2004, a quarter of doctors made money promoting a drug (Campbell et al., 2007); it should be borne in mind that this not 25% of university researchers, but rather standard, working, patient-facing doctors. Indeed, many of the doctors hired in this way were not contracted because they were especially persuasive or knowledgeable, but because they were high-volume prescribers. That is, they were ostensibly hired to talk about a drug to other doctors, but, unbeknownst to the speaker, they were selling themselves.

A single talk for a local doctor may sometimes only add a few hundred dollars to a paycheck, but such payments vary considerably. There are over 25 elite psychiatrists in the UK alone taking home in excess of $200,000 in addition to what them make as professors (Healy, 2012, p. 222). In Minnesota, where sunshine laws were first passed, over 100 doctors received more than $100,000 from pharmaceutical companies (Bass, 2008, p. 225). All told, in 2003, pharmaceutical companies spent about $18.5 billion marketing to doctors, or roughly $30,000/doctor/year (Kassirer, 2005, p. 77). This figure does not include other means companies use to influence doctors or funnel money to them, such as the "efficacy study" Forest used to get doctors accustomed to writing prescriptions for Celexa (cf. Avorn, 2004, pp. 210–211).

Much of the commercialism is cloaked in respectable sounding language. When doctors are paid to speak, they are called key opinion leaders (Elliot, 2010, pp. 75–108).[8] Detailers are of course salesmen, but the transaction is veiled. The doctor never buys anything from the detailer and so the encounter is somewhere between a quick chat with an acquaintance and mini-educational lecture. Detailer-turned-medical anthropologist Michael Oldani emphasized that in order to be effective, the "pharmaceutical gift-economy" transpires under the pretense that everything is being done in the best interest of the patient (Oldani, 2004). However, the CEO of one ad agency puts this more simply:

> Very often doctors are more influenced by what other doctors say than what pharmaceutical companies have to say. So companies work through medical education companies to have doctors that support their products talk about their products in a favorable way. That's called medical education. (Torre, as quoted in Kassirer, 2005, p. 92)

1.6 Commodification of knowledge and seeding trials: A threat to the reliability of medical research

For our purposes, it is crucial to make a distinction between private- and public-oriented research. Here, private-oriented research refers to the early work in drug discovery, such as synthesizing and manipulating chemical structures, drug screening, and determining how a drug is metabolized; in short, the basic work that goes into drug discovery. This process is generally meant for private consumption within a company. As it forms the basis of choosing which chemical is most likely to be successfully turned into a drug and there is a high cost for failure, there is an internal pressure to generate reliable knowledge. On the other hand, the intended audience of public-oriented research is people outside the company, such as regulatory officials, medical doctors, and academics. Because the function of this research is to create and foster a successful product, there is an internal pressure to produce favorable

[8] The industry line on such relations is that by funding experts they are ensuring that doctors stay up to date on the most effective treatments. If companies only promoted useful drugs, this justification may have some foundation. As the former vice-president of Pfizer sees it, however, such claims are ludicrous: "When drug companies sponsor medical opinion leaders, when they send them on trips across the country, or give grants for their research or to their universities, no one should believe it is out of charity. It is done for one reason, and that is to sell more drugs" (Rost, 2006, p. 189).

results. While the distinction is important in the context of understanding the scientific role and contribution of pharmaceutical companies, it is the latter type of research that is seen as problematic.

The most direct way in which research has been commercialized is in the conduct of trials where the sales targets are the doctors ostensibly conducting the study (Angell, 2004). A seeding trial is one in which doctors are paid to "enroll" patients, but where there is no epistemic purpose. As the actual goal is to get the participating doctor accustomed to prescribing the company's drug, these "studies" are designed and run by the drug manufacturer's marketing department. Sometimes the veil of scientific legitimacy is not even donned, as when Pfizer's marketing department conducted a trial under the project name CRAM (Central Research Assists Marketing) (Healy, 2004).[9]

However, even trials designed to produce data are driven by economic considerations. Industry now funds a considerable amount of medical research, both in-house and by awarding research contracts. The conduct of research in the context of heavy industry funding is a central concern for nearly every author concerned about pharmageddon.

A typical conception of scientific research is that an experiment is constructed to test a hypothesis or to explore some unknown aspect of nature. As Mayo and Spanos (2010) stated, as opposed to testing some grand theory, most research has

> the goal of 'finding things out,' the researcher in [Mayo and Spanos'] view is driven to create novel effects and to compel the replacement of hypotheses and theories with ones that not only give us a more correct understanding about more domains, but also teach us where our earlier understandings are in error. (p. 85)

This ethos stands in stark contrast to the recommendation of veteran drug developer Lawrence Friedhoff (2009), who proposed the following "rule of thumb:"

> Don't investigate questions unless you can accept all the realistically possible answers. For example, do not undertake a study designed to show your drug is superior to a competitor unless you are fairly sure that the study will validate your hypothesis of superiority. (p. 97)

[9] It should be noted that when companies provide figures for how much they spend on research, these "studies" are included in the total (Angell, 2004).

Industry-funded studies bear far more resemblance to the purchasing of facts than processes of discovery. As Friedhoff suggested, it is often the same uncertainty that drives veritistically oriented researchers to pose a question that causes industry to be cautious about seeking the answer. This is not to suggest that the results of industry-funded studies are necessarily false. For example, when the antidepressants Wellbutrin and Zoloft were first released, their manufacturers funded studies to demonstrate the real prevalence of drug-induced sexual dysfunction in Prozac; a fact that Prozac had been designing studies to ignore (Glennmullen, 2000).[10] One might think of the value of conducting research as follows: (probability of obtaining the finding) x (the increase in sales) – (probability of the alternative hypothesis) x (the decrease in sales) – (cost of the study). One can almost see this calculation being made by drug companies in the following conversation reported by Harvard professor Jerry Avorn, in which he sought funding from the National Pharmaceutical Council (NPC) for a study that would potentially bolster the case for increased government spending on pharmaceuticals:

> *Avorn*: We are trying to understand the problems that patients run into when their drug coverage is taken away.
>
> *NPC:* What do you expect the results to be?
>
> *Avorn*: We can't be sure, but clinically I'd expect that if you take away heart medication from cardiac patients, and insulin from diabetics, and inhalers from asthmatics, then bad things will happen.
>
> *NPC:* Is there a chance you might find no difference in the results?
>
> *Avorn*: Of course, there is that chance; that's why we do the study. But I don't think it's likely.
>
> *NPC*: Would you publish your results either way, even if there was no difference in health care outcomes?
>
> *Avorn*: Yes, we would.
>
> *NPC*: Let me get back to you on this. [The study was declined]. (Avorn, 2004, pp. 194–195)

[10] To forestall the impression that such tactics are confined to psychiatry, see Fugh-Berman (2005). Fugh-Berman was asked by AstraZeneca to be the nominal author of a study published several years before their blood thinner Exanta came out. The ghostwritten article highlighted the possibility that their prospective competitor's drug had harmful interactions with common herbal remedies. The interchange was a revelation for the Georgetown professor: "The thing I didn't know was that companies funded articles to kill a small part of the market. I also didn't understand how far in advance of a drug launch they started seeding the literature" (Fugh-Berman, 2005, p. 548).

As we saw in the case of Celexa for pediatric depression, even a single positive study can be leveraged to generate massive sales. On the other hand, a negative trial may not just result in poorly spent funds, it could "undermine the commercial profile" of a drug.

When viewing things merely in terms of their effect on the corporate bottom-line, a number of practices become completely rational. For example, there are several means of increasing the probability of a positive finding: manipulation of research design (Safer, 2002); hiring a ghostwriter to ensure the results are construed in a favorable manner (Fugh-Berman, 2005; Golden, Parochka, & Overstreet, 2002; Rennie & Flannigan, 1994; Ross et al., 2008); selecting researchers known to hold sympathetic views (see Chapter 2); and so on. Similarly, another way to increase the expected returns is to reduce or eliminate the probability of negative studies becoming public. This can be done in a number of ways, such as having researchers sign nondisclosure contracts (Schulman et al., 2002); threatening to sue researchers/journals if they publish negative findings (Fox, 2010; Gibson, 2006; Moynihan, 2010); conducting research in-house; or contracting private research companies instead of academics (Angel, 2004).

This latter option has become extremely prevalent in the past 20 years, as it significantly reduces the cost (as defined above) of conducting research. Contract research organizations (CROs) emerged in the 1980s, became a billion dollar industry by 1990, and grossed three and a half times that by 2000 (Whitaker, 2002, p. 263). Whereas in 1990 80% of industry research was conducted in the university setting, by 2000 the figure had dropped to less than 40% (Angell, 2004). Faced with competition from CROs, university researchers began to adopt a partnership model, in which institutions have given up control of trial design, access to data, and/or the freedom to publish (Schulman et al., 2002).

When it comes to such trials, the primary concern is not fraud, but truthful and misleading answers to carefully crafted questions.[11] Commensurate with such concerns are a number of meta-analyses that have found a large correlation between positive results and industry funding (Bekelman, Li , & Gross, 2003; Bhandari et al., 2004; Ridker & Torres, 2006). Rochon et al. (1994) examined comparison trials funded by manufacturers of nonsteroidal anti-inflammatory drugs for arthritis and found that 56/56 studies concluded that the funder's product was as good or better than the comparison drug. In essence, when drug A is compared with drug B, the better drug is determined by the funder of the research. While this was particularly egregious, it is estimated that between 89% and 98% of trials yield results favorable to the company that has funded the research (Cho & Bero, 1996; Davidson, 1986).

A related, and perhaps more troubling, concern relates to extremely important questions that have no fiscal upside. Where doctors already prescribe a drug for long-term treatment, there are only economic downsides to conducting protracted and costly studies to determine its long-term effects. For example, regarding the mounting concerns that long-term use of antidepressants has harmful neurological effects when paired with the dearth of rigorous studies addressing the question, Columbia psychiatrist Donald Kline decried,

> the industry is not interested, the NIMH [National Institute of Mental Health] is not interested, and the FDA is not interested. Nobody is interested. What balks everybody is that it [a long-term study] would be expensive and difficult. I think

[11] I do not mean to discount pernicious effects of the "graphite method." Fraudulent data provides the most serious and most difficult-to-detect bias. For example, in the published version of Paxil study 329, patient 70 was listed as a 12-year old boy that had discontinued imipramine (an older antidepressant) because it caused tachycardia. A letter from the lead researcher at Brown University to the institutional review board shows that patient 70 was in fact a teenage girl that had discontinued Paxil after being hospitalized for attempting suicide. At least two other suicide attempts were coded as if the patient had voluntarily withdrawn from the study and stopped taking Paxil. This fraud was not detected by internal checks for scientific integrity, but because the trial was investigated by the New York Attorney General (Bass, 2008; cf. Whitaker, 1998). Even the FDA reviewers, who are generally the most scrutinizing audience of randomized clinical trials, generally cannot assess the accuracy of data and so resort to evaluating its consistency in the hope that if it has been falsified this has been done so carelessly (Friedhoff, 2009). However, for every fraudulent study there are likely hundreds of accurate, albeit misleading, results.

> industry is concerned about the possibility of finding long-term risks. (Kline, as quoted in Glenmullen, 2000, p. 105)

The FDA does not require trials of a drug commensurate with their actual use, and doctors use drugs for durations far beyond the time where research has demonstrated efficacy. Thus, from a financial perspective, funding long-term studies involves immense costs and no potential upside.

With industry funding the majority of research, the medical community is left with an overly optimistic impression on a set of research questions that are already slanted towards yielding positive results. Research areas that can only produce knowledge of a lack of efficacy or harm are neglected altogether. It hardly needs to be noted that from an epistemic perspective a proper balancing of risks and benefits is next to impossible in such circumstances. Perhaps it would be unwise to withhold drugs that demonstrate short-term benefits from the general population for years while longer trials are assessed; however, failing to conduct such studies at all has the potential to lead to tragedy. Moreover, training doctors to base treatment decisions on clinical trials has only increased the importance of favorable results to a company's fiscal health. Axis Healthcare Communication, a medical communication company, offers services such as publication planning and medical education services in order to, as Axis put it, "brand the science." Similarly, the CRO Quintiles promises to "Generate the right data, for real results" and encourages pharmaceutical companies to avail of its "Scientific guidance … to overcome any obstacle" (Quintiles, 2014).

1.7 Pharmageddon revisited

Concern over the state of medical research ebbs and flows. On the heels of a great tragedy (e.g., Elixir Sulfanilamide, thalidomide, pediatric use of antidepressants) the medical community tends to stop and reflect on what has gone wrong. It is at such times that the focus is less on methodology and more on the unseemly influences on medical judgments. The emphasis is less on truth, and more on what experts believe to be true, or what Merton (1942) called certified knowledge. The pressing question is

always, how did certified knowledge stray so far from what doctors were in a position to know? In such a reflective state, the community reexamines its practices around nosology, regulation, publication, education, and research because these practices are the means of certification (and thus of manipulation). The past decade has seen a spate of such reflections.

When close attention is paid to the way in which medical knowledge is certified, there is inevitably and understandably a shocked horror and condemnation at the violations of norms of science. There is also typically a false belief that the recent awareness of commercialism is due to a new emergence of commercialism. By comparing the unvarnished present to the gilded past, authors are misled into thinking that the medical community is on the verge of pharmageddon, an epistemic hell in which the official organs of knowledge cannot and should not be trusted.

I will argue that while such concerns are overblown, the issues raised are not groundless. Medical knowledge has inherent moral import and misleading doctors inevitably leads to inflated costs, unnecessary suffering, and untimely deaths. People should be outraged. Further, the simple fact that we are not in epistemic hell does not imply that these threats are not real. The central argument of the next chapter will be that because these influences are not taken into account in current epistemologies, philosophers draw the wrong lessons from case histories.

To accept that doctors will be misled in matters of life and death is to lose sight of the fact that today, as in the past, there have been dedicated groups of professionals who seek epistemically reliable grounds for medical treatment. There is no reason to believe that reformers today are any less capable than those of the past. Previous reformers have not purged the industry of commercial distortions, but neither have their efforts been in vain. This dissertation uncovers how to think about medical epistemology given that both commercial and veritistic impulses remain driving forces in medical epistemology.

1.8 Plan of the dissertation

In Chapter 2 I examine philosophical accounts that attempt to ignore the problems outlined above. I argue that there are some questions that can indeed be answered without including the effects of commercialism. On the other hand, philosophers' concerns often go beyond this domain. I show that ignoring the commercial drivers of medical research leads philosophers who ignore them to draw the wrong lessons from cases they take as paradigmatic evidence for their views.

In Chapters 3 and 4, I propose my own account of how to integrate both veritistic and commercial drivers of medicine into a single coherent view of medical epistemology. Specifically, I argue that medical epistemology should be seen as an example of an asymmetric arms race. Chapter 3 is dedicated to articulating a general theory of asymmetric arms races, while Chapter 4 reviews the early history of medical regulation to demonstrate that medical epistemology is, and always has been, an asymmetric arms race.

In the final three chapters I discuss the difference it makes to view medicine as an asymmetric arms race. In Chapter 5 I expose a problem that exists in the standard way in which we evaluate the efficacy of drugs, and suggest a solution to this issue. Though the solution is novel, the problem has been an open secret in drug testing. The fact that a known problem has remained unaddressed is a testament to the fact that satisfying the regulatory requirements of the FDA, rather than methodological rigor, is the primary goal of significant portions of those who conduct medical research.

In Chapter 6 I consider a formal model of epistemic communities which presumes that all agents are truth-seeking. After recognizing medical epistemology as an arms race, it becomes clear that such a model fails to include the effects of the commercial imperatives of medical research. I show that the canonical results of this model fail to obtain when even one of the agents involved is motivated by convincing the group of a favored position. After exploring the dynamics of such a network, I consider how agents might attempt to mitigate the influence of biased agents.

Finally, in Chapter 7 I consider what implications the arms race picture has for understanding medical epistemology. In particular, I argue earlier chapters show that medical epistemology must be a social epistemology if it is to address issues in practice. The complications caused by commercial imperatives are an entrenched part of medical research, not some recent aberration. Moreover, because of the effect of antagonistic forces, one cannot equate epistemic reliability to methodological rigor. Finally, I argue that robustness (i.e., resistance to manipulation) is the proper standard by which to judge future reforms.

CHAPTER 2

Philosophers on Drugs

The Limitations of Friction-free Epistemology

2.0 Preface

In Chapter 1 I examined a number of ways in which medical research has been perverted to serve commercial ends. Traditionally, such exposés have been considered the purview of ethicists; however, epistemologists and philosophers of science have increasingly come to realize that the evaluation of evidence in areas with potential impacts on public policy and the public good is fundamentally entwined with issues of deep ethical significance (e.g., Douglas, 2009; Elliott, 2011; Kitcher, 2001, 2011; Kourany, 2011). Nevertheless, the dominant approach to medical epistemology is one that prevents the important ethical issue from even arising or being visible, because it presumes that inquiry is being conducted by a community of truth-seekers acting in good faith. The most recent standard-bearer of this narrative has been philosopher Jeremy Howick (2012), but many others (e.g., Cartwright, Mayo, and Worrall) have put forward their own views on the nature of medical evidence. These philosophers have tended to conceive of medical epistemology as an effort to identify and deploy the most rigorous methods in the most effective way. Based on this framework, philosophic disputes have centered around what levels of absolute and comparative confirmational significance should be assigned to particular types of evidence. Most crucially, the abovementioned authors largely ignore or abstract away from questions about who is actually making use of these methods, or how.

In this chapter, I will examine previous accounts that have either ignored the effects of pervasive industry funding (e.g., Cartwright, Mayo, and Worrall) or largely brushed it to the side as a "sociological problem" (e.g., Howick). Specifically both Cartwright and Worrall discuss "ideal randomized clinical trials (RCTs)" and use these idealizations to discuss the limits of inference. Any account that abstracts away worldly complications to consider ideal experiments will be referred to here as *friction-free epistemology.* To an extent, this is all well and good. Abstractions are an excellent way to ignore messy details and lay bare some of the underlying structure of the limits of inference. In such cases, *friction-free epistemology* is entirely appropriate, as one is essentially asking: In the best of all

possible worlds, what am I entitled to infer, given evidence of such and such? Idealizations are fine for ideal worlds.

Yet there are further questions that philosophers take themselves to be addressing. These are questions such as: Was the meta-analysis performed by Patricia Crowley (1981) sufficient evidence to make treatment of infants with corticosteroids standard practice, despite the fact that experts judged corticosteroids to provide no benefit (Howick, 2012, p. 161)? In general, Howick framed his book as "an evaluation of the EBM [evidence-based medicine] view of what counts as 'good evidence'" (Howick, 2012, p. 24). Similarly, Worrall (2007) viewed medical evidence as "a new area where philosophers of science could have enormous impact—both intellectual and (very unusually) *practical*" (p. 981, emphasis added). The question here is what actual doctors should make of actual evidence. For questions such as these, friction-free epistemology works only as well as the actual world approximates the ideal.

It is my contention that the way medical science works, and thus the epistemic reliability of science in this area, is far from ideal. Specifically, I propose that any plausible account of medical knowledge must incorporate the fundamental antagonism between science and commerce. I will look at the philosophical literature on causal reasoning in medicine as an example of a philosophical debate that has been going on the abstract (Sections 2.1 and 2.2). I will then examine the details of a case that EBM proponents take to pertain to the dangers of causal reasoning (Section 2.3). I will argue that philosophers have misidentified the primary threats to medical knowledge (Sections 2.4 and 2.5), and then reject a modified position that industry funding can be considered one of many possible confounders (Section 2.6). I will conclude by drawing a distinction between *error* and *manipulation* (Section 2.7). Given that philosophers of science are primarily concerned with how medical research can avoid errors, while the main threat to medical knowledge is manipulation, I contend that philosophers have been focused on the wrong issues, have failed to address precisely those questions that should

matter most to our evaluation of evidential warrant, and as a result obscured some of the most pressing questions in medical epistemology.

2.1 Philosophers on drugs: Do doctors need causal theories to prescribe rationally?

Cartwright (2007, 2009, 2010, 2011) has been building the case that EBM has come to rely too heavily on RCTs. She noted that proponents of EBM often favor an atheoretical account of efficacy—a desire to let results stand alone without a narrative overlay about why such results have occurred. For example, the Cochrane collaboration, which has heavily influenced EBM, privileges highly controlled trials on the assumption that causes identified in highly controlled settings will continue to operate in the field. Many theories relevant to RCTs are controversial, ill-formed, or poorly supported; thus, some proponents of RCTs have argued that these theories should be eschewed. In practice, significantly outperforming a placebo in an RCT is sufficient to gain approval from the FDA, regardless of whether there is a well-supported theoretical account of the treatment's efficacy.

But as Cartwright (2007, 2009, 2010, 2011) has argued, an atheoretical account will not license generalization from the experiment to the conclusion that the treatment will work in practice, which is the presumptive purpose of the experiment. The result of an RCT concerns the difference between two group means, but what is needed is a justification to act in a particular instance outside the experimental setting. Cartwright argued that this requires three things: (1) reason to believe that the effect observed in the RCT is the product of a capacity that will endure outside the experimental setting; (2) reason to believe that the proper causal structure has been identified; and (3) an account of what it means for the cause to *contribute* in new situations (some description of how the cause will interact with other causes relevant to the non-experimental setting). In short, there is need for a theoretical account of what has been observed.

The problem with causal mechanisms is that they are both too much and too little. There are many cases in which we have good evidence that something works and no causal account of why, such as Edward Jenner's development of vaccination for smallpox 60 years before the germ theory of disease. The cancer drug cisplatin, which is widely regarded as one of the best antitumor drugs in the modern armamentarium, is a modern example. While there are working hypotheses on the mechanism of cisplatin, decades after its approval "the molecular details of how it causes cells to die are largely unknown" (Fink & Howell, 2000, p. 149). It is not that having a causal account would fail to aid rational use of treatments, it is just that causal mechanisms are clearly not the only warrant for employing a validated treatment in an individual case. Moreover, as the following example shows, having a causal account can sometimes be detrimental.

2.2 The danger of relying on causal theories: A philosopher's account of a medical disaster

In the early 1980s, people started dying in unexpected ways. At first only a few doctors noticed the trend, but soon people all around the world were dying en masse. Faced with such dire circumstances, Ronald Reagan's National Institute of Health (NIH) stepped in and orchestrated a multicenter RCT employing thousands of medical professionals in 26 cities across the world. The trial, begun in 1987, became a flash point as ethicists debated the acceptability of giving half of the patients a placebo in lieu of treatment. By the late 1980s, tens of thousands of people were dying every year; estimates of the death toll for the decade range from 240,000 to 700,000 in the United States alone. It was loss of life on the order of American casualties in World War II. The cause of death was not AIDS (roughly 40,000 deaths in the US from 1980–1989), but the result of using class-I antiarrhythmic drugs.

The details of this story are crucial because the case serves as a paradigm example in the argument that RCTs, not causal accounts, should be the basis of treatment decisions. Paradigm cases such as this are intended to illustrate a problem, explain why it occurred, and identify what can be done

to prevent future tragedies. In this section I will provide the account as given by EBM proponents (Evans, Thorton, & Chalmers, 2006; Howick, 2012; Strauss et al., 2011), and in Section 2.3 I will present a fuller narrative that illustrates how the standard account misidentifies the causes of, and thus the solutions to, the arrhythmia disaster. After the history has been established, I will argue in Section 2.4 and beyond that current philosophical accounts fail to address central aspects of medical epistemology.

In order to provide context on the EBM narrative, I quote Howick at some length:

> Consider a rather famous case of mechanistic evidence involving antiarrhythmic drugs to reduce mortality in patients who had suffered myocardial infarction (heart attack). Myocardial infarction often damages the muscle and electrical system in the heart, leaving it susceptible to arrhythmias. A common type of arrhythmia, ventricular extra beats (VEBs) occurs when the left ventricle contracts before it has time to fill completely. The heart then fails to pump sufficient blood. Without treatment, lung, brain, and kidney damage ensues. Worse, VEBs can also degenerate into ventricular fibrillation in the absence of electrical shock. Large-scale epidemiological studies suggested that between 25 and 50% of sudden cardiac deaths were associated with arrhythmias. Based on this understanding of the underlying mechanisms, several drugs were developed and found to be successful for regulating VEBs. The drugs became widely prescribed in the belief they would reduce cardiac deaths. (Howick, 2012, pp. 126–127)

Given the widespread use of antiarrhythmic drugs, the NIH convened a study (CAST I) to test the arrhythmia suppression hypothesis. Midway through the study an interim advisory committee was convened, and because the data were highly significant and unlikely to change, it was deemed unethical to continue the trial. When the advisory group was unblinded, it was discovered that the group with more deaths was the treatment group, not the placebo group. At the time the trial was halted, 56 of the experimental patients had died (compared to 22 on placebo). The FDA was called and the *NEJM* expedited a manuscript of the study showing how dangerous doctors' assumptions had been. The results were subsequently confirmed in a second study (CAST II). [12]

[12] These figures do not include a large number of patients who died in the first two weeks of the trial, when all patients were on an antiarrhythmic drug to establish the effectiveness of VEB suppression. Of the first 91 deaths, 41 occurred during the two-week titration phase of the trial (during which there was no placebo group for comparison). These deaths were not included in the analysis. In CAST II, which used another class I antiarrhythmic drug, 17 patients died during titration compared to three on placebo before this trial was also stopped.

Howick (2012) contended that such cases show the frailty of relying on mechanical reasoning:

> The mechanism(s) involved in antiarrhythmic drug action might include swallowing and gastric emptying, metabolizing circulatory and binding mechanisms. The mechanisms involved in getting the orally administered drug to its pharmacological targets on the cells are relatively (but not completely) well understood and referred to in the medical literature as ADME (mechanisms for absorption, distribution, metabolism, and excretion). Once the drug reaches its cellular target, antiarrhythmic drugs reduce the frequency of VEBs by modifying the heart's electro-chemical mechanism. Finally, a reduction in VEBs (allegedly) reduces the risk of sudden death… Mechanistic reasoning involves *inferring* from knowledge of the mechanism(s) to the claim that an intervention has its putative effects. I might know about the ADME, heart and brain mechanisms, but to move from there to the claim that antiarrhythmic drugs will reduce mortality, I must be able to predict what happens to each of the mechanisms under intervention. (p.127f)

This poses a challenge to Cartwright's (2007, 2009, 2010, 2011) position by pointing out that we rarely have a sufficient account of a drug to determine how it will interact with biological mechanisms. This complexity makes causal reasoning based on mechanisms an unreliable source of evidence. Of course, we do rely on causal reasoning when we depend on surrogate endpoints[13] (such as suppressing VEBs) to evaluate drugs, but these can be sources of significant tragedies, as seen in the arrhythmia case. The problem arises when "some but not all of the relevant mechanisms involved in mechanistic reasoning are not based on sound evidence … [in this case] the impact of the drugs on mortality was not based on sound evidence" (p. 138). If Howick is correct, then it is not so much that mechanistic reasoning was relied on, but that the mechanical reasoning for the patient-relevant outcome (mortality) was not "high quality."

This analysis identifies the cause of the tragedy as the reliance on low-quality mechanical reasoning. Howick suggested that if the doctors had simply paid closer attention to the epistemic

---

Antecedently, the drug studied in CAST II was thought to be the "most benign" member of the class. Since the initial dose given was low and high doses increase risk, the results "may be considered a minimal estimate of the risk of the initiation of drug therapy" (CAST II, 1992, p. 230).

grounds of the arrhythmia suppression hypothesis, they would have seen that it was not fully warranted. Because it was in fact partially warranted (the drugs did stop VEBs), it had the "aura of acceptability, which in turn [led] to more prolific use of a harmful treatment" (pp. 139–140). The solution to the problem is to educate doctors not to trust low-quality mechanistic reasoning, and in most cases to insist on RCTs of clinically relevant outcomes before a treatment is widely used. It would seem that the only thing left do is to specify "high-quality" mechanistic reasoning and such calamities can be avoided in the future.

Howick's narrative on the arrhythmia case is oversimplified. Delving into the case in greater detail will illustrate that from the very beginning the commercialization of medical knowledge served as the wellspring that fed the widespread success of the deadly medicine. The fuller story underscores why medical epistemology cannot be separated from the "sociological problems" caused by commercial imperatives. A corollary is that philosophers have misidentified the dominant forces threatening a sound medical epistemology. It bears repeating that such problems occur throughout medicine (Angell, 2004; Avorn 2004; Brody, 2007; Elliot, 2010; Healy, 2012; Kassirer, 2005; Krimsky, 2003). The point here is that because such problems are pervasive, they can be identified even in cases that are selected as paradigm examples for other accounts. The fuller account of the arrhythmia disaster will take us into the details of medical history, but the philosophical payoff depends on understanding how the account given by EBM proponents falls short.

2.3 This time with friction: How commercial forces contributed to the antiarrhythmic drug disaster [14]

> *Everything in war is very simple, but the simplest thing is difficult. The difficulties accumulate and end by producing a kind of friction that is*

[14] This account draws heavily on the book *Deadly Medicine* (Moore, 1995), which is often referenced by EBM advocates as their source of information on arrhythmias. I will indicate where my account differs from Moore's rendition in footnotes.

> *inconceivable unless one has experienced war... Friction, as we choose to call it, is the force that makes the apparently easy so difficult.*
> (von Clausewitz, 1832/1989, pp. 119–121)

The account given by EBM proponents locates the source of the arrhythmia disaster in the reliance on causal reasoning. This allows it to serve as a case example of the unreliability of causal reasoning and as an argument against philosophers who stress the necessity of identifying mechanisms in warranting treatment (Cartwright & Hardie, 2012; Russo & Williams, 2007). In this section I will develop a single case and show how the commercialization of medical evidence (disease-mongering, a weak FDA, an imbalance in the confluence of interests, marketing to doctors, and the commercialization of medical trials) each contributed to the arrhythmia disaster. I claim that it is these factors, and not a reliance on poorly supported causal reasoning, that is at the heart of the problem. Thus, to the extent that philosophers take themselves to be addressing threats to actual inferential practices, they are addressing the wrong issues.

The story of the arrhythmia case begins back in 1978 when Bernard Lown gave his keynote address at the annual gathering of America's cardiologists (Lown, 1979). Lown's early work had led to the invention of the defibrillator, but patients had to be reached within minutes before organ failure and brain damage began. If patients were reached shortly after ventricular fibrillation, they often led long and healthy lives. In his speech, Lown crystalized the arrhythmia suppression hypothesis he had been working on for over a decade. Several studies showed that VEBs were correlated with increased risk of death, and for Lown this was key for his prophylactic campaign.[15] Several decades of research suggested that VEBs precipitated cardiac arrest, even after controlling for underlying structural damage to the heart.[16]

[15] The various terms used in the medical literature for this phenomenon were "premature ventricular complexes," "ventricular premature complexes," and "extrasystoles." I have chosen to stick with Howick's acronym for the sake of simplicity.

[16] Moore (1995) suggested that the cardiac suppression hypothesis was only based on simple correlation, and that it was equally plausible that heart arrhythmias and subsequent cardiac arrests had a common cause, but were not

At the time Lown proposed the arrhythmia suppression hypothesis, antiarrhythmic drugs already existed to treat sustained rapid beat. In these cases, the heart beats too rapidly to pump blood effectively, and if the rapid beats continue, the patient typically dies within a matter of days. In contrast, people with VEBs have a few extra beats sporadically, but the majority of the time their heart beats normally. Because these latter patients are typically unaffected, it was not until the advent of sophisticated heart monitoring that VEBs were even identified. Given the known risks of antiarrhythmic drugs, Lown repeatedly emphasized in his speech that VEBs, even when they are frequent, "require no treatment at all other than the physician's affirmation of their ubiquity and benignity. Therapy is needed in only a minority of patients, who usually have ischemic heart disease and a life-threatening or symptomatically disabling arrhythmia" (Lown, 1979, p. 321).

Lown separated arrhythmias into five different grades and contended that it was only for the two most severe grades that the benefit of treatment outweighed the risks (Graboys et al., 1982; Lown, 1979). This concern was born out by studies that showed the troubling propensity of drugs that precipitate arrhythmogenic effects (i.e., that worsen arrhythmias). In roughly 11% of patients, the drugs could precipitate cardiac arrest that was unusually hard to reverse and unlike anything doctors had seen before (Velebit et al., 1982; Winkle et al., 1981). However, the assessment that antiarrhythmic drugs are dangerous and therefore should be restricted to a small subclass of patients was actively opposed by industry-supported academics. These researchers argued that a far greater patient pool stood to

---

themselves causally related: "Every doctor has been taught that an association between two events does not prove a causal link. In this case it was equally plausible that premature beats were nothing more than a telltale indicator of underlying permanent damage" (p. 49). Howick (2012) repeated this claim: "The available evidence suggested an epidemiologic link between VEBs and mortality, but association is not causation. Moreover, even at the time there were good reasons [he did not specify what these were] to believe that after myocardial infarction, the heart is damaged in a way that both causes VEBs and raises the risk of sudden death" (p. 138). The supposition that the arrhythmia suppression hypothesis confused correlation with causation is simply untrue. For example, in the very speech in which Lown described his theory, he addressed the topic as follows: "It may be argued cogently that prognostic implications are not determined by the ventricular premature complex [VEBs] but by the extent of cardiac disease because the grade of ectopic activity is largely an expression of the severity of the disease. A corollary inference is that the attempt to control ventricular arrhythmia is futile because the ultimate outcome is determined by the extent of heart disease. A recent study of Schulze et al. contradicts such a conclusion" (Lown, 1979, p. 316).

benefit from the prophylactic effects of antiarrhythmic drugs (Anderson, Stewart, & Crevey, 1984; cf. Bigger, 1984). They found the prevalence of arrhythmogenic effects to be "low and rarely lethal." Similarly, they discounted drug-induced changes in heartbeats that were generally thought to be dangerous. For example, in evaluating the antiarrythmic drug Tambacor (1984) such alterations should "not necessarily be a firm indication to discontinue flecainide [Tambocor] when the patient otherwise has a successful therapeutic response" (p. 93B).

The variation between the two viewpoints made an enormous difference to the patient population that would be treated. If Lown was right that only severe cases merited treatment, it would cut out nearly 85% of the population that manufacturers had intended to market the drugs to. Though the FDA originally intended to approve the labeling of antiarrhythmic drugs only for the more restricted population, the 3M corporation persuaded the FDA to approve the wider indication.[17] Internal memos from 3M leave little doubt as to the reasons why the FDA's initial labeling was unacceptable: "This labeling is most detrimental to our market success and is indeed the worst case scenario … In order to meet financial projections for the drug it is imperative to have our original label approved" (as quoted in Moore, 1995, p. 145). With this success, 3M made sure doctors knew about the risks of VEBs by promoting the expansive version of the arrhythmia suppression hypothesis. Because 3M was not the only company that would market antiarrhythmic drugs, and since a good part of selling the drugs was promoting VEBs as a treatable condition, cardiologists were repeatedly informed about the dangers of VEBs. The expansion of a diagnostic category in order to increase the profitability of a drug is clear example of *disease-mongering.*

The prospect of a huge untapped market for pharmaceutical intervention led to a number of companies patenting and commencing to test pharmacologically similar drugs. Given that these drugs

[17] The labeling of a drug indicates the FDA-approved usage for a drug. Doctors may use drugs for a non-approved condition, but companies can only promote their drug for the indication approved by the FDA.

ultimately caused increased mortality, the importance of VEBs as a surrogate endpoint is clear in retrospect. While the FDA requires that a company must provide evidence of efficacy from adequate and well-controlled trials in order to be approved, the criteria for efficacy are determined by expert consensus. When a variable of interest is difficult to measure or would take years for a trial to establish, experts sometimes decide to use a surrogate for the variable because of its purported relation to the clinically relevant outcome. For example, in AIDS trials T-cell counts were used as a surrogate endpoint for mortality; this was based on an assumption that there was a causal relation between low T-cell count and death, rather than being because T-cell count is an intrinsically clinically important variable. In this case, the FDA relied on the expert consensus that control of VEBs was a suitable surrogate endpoint for mortality, and this consensus was reached at a conference convened by the very researchers whom pharmaceutical companies were paying to run trials of their antiarrhythmic drugs. This was the first of a number of small decisions on which there was scientific disagreement that was ultimately decided in the manufacturers' favor.

Contrary to the standard account offered above, the researchers were well aware that the arrhythmia suppression hypothesis was not definitive. In a panel discussion of 3M's drug Tambocor, Betram Pitt (an academic researcher funded by 3M) reminded the panel that the drug had known arrhythmogenic effects and asked: "What happens if it turns out, when the studies are done, that far more of the people died who got this drug?" To which Joel Morganroth (a rising star in the arrhythmia world, receiving research grants from 3M as well as at least 12 manufacturers with a vested interest in VEBs being accepted) replied: "That is a decision to be made at the time those studies are done" (as quoted in Moore, p. 134).

In contrast to the standard account, research physicians were aware of the large number of deaths that had occurred while patients were on Tambocor. The drug was approved because when the industry-funded experts went through and examined deaths that occurred in medicated patients, they

essentially assigned blame to the drug only when no other cause suggested itself. This and a series of other small judgments ultimately led industry-funded researchers to provide a far different estimate of the risk/benefit ratio compared to independent researchers. In many cases, it is difficult to determine whether panel members adopted commercially beneficial positions because they worked for the industry, or worked for the industry because they held a commercially beneficial position. Yet, had the FDA convened an independent panel to evaluate whether suppression of VEBs was an acceptable surrogate endpoint, the outcome may well have been completely different.[18]

This is because there were several critics of the more expansive version of the arrhythmia suppression hypothesis (including Lown himself). One particular example is instructive here. Among the first researchers to test Bristol-Myers Squibb's antiarrhythmic Enkaid (encainide) was Stanford cardiologist Roger Winkle. During the initial trials, a patient on Enkaid developed a rapid heartbeat that was extremely difficult to reverse and did not respond to standard interventions. Concerned, Winkle reviewed the 140 other cases he had treated with Enkaid and found 10 other serious incidents, some of which precipitated the death of the patient. Winkle attempted to publish his findings in *Circulation*, but the article was rejected. It was then rejected twice more from other journals.[19]

Fortunately, the Stanford cardiologist knew the editor of the *American Heart Journal,* which ultimately published the article, but without such connections, the prospects for publication would have been uncertain. Indeed, it is now known that some early trials identifying sudden death in patients treated with antiarrhythmic drugs went unpublished (e.g., Cowley et al., 1993). While there is no

[18] In addition to those mentioned above, the FDA Cardiovascular and Renal Drugs Advisory Committee had several members with ties to companies whose drugs were up for approval, including Raymond Woosley from Vanderbilt and Edward Pritchett from Duke.

[19] It is difficult to determine the exact causes of the rejection and whether bias (conscious or unconscious) played a role in any individual case. Nevertheless, some of the referee comments that Winkle received were difficult to understand. For example, his article was rejected from *Circulation* for the following reason: "because the literature does not at present contain an overall description of the antiarrhythmic efficacy of Enkaid, it seems somewhat inappropriate to include a separate article about this specific adverse side-effect" (as cited in Moore, 1995, p. 66). This explanation is odd given the existence of four separate studies describing the antiarrhythmic efficacy of Enkaid, all of which were cited in Winkle's manuscript (1981).

smoking gun that journal editors consciously refused to publish damaging information about one of their advertising clients, the difficulty that Winkle encountered underscores the additional hurdles commercially damaging evidence encounters.[20]

Meanwhile, 3M was able to publish the same dosing trial six times, including in three top journals (*The American Journal of Cardiology, Circulation, NEJM,*). The trial had been designed and analyzed by 3M, but each time the article had different lead authors from academia. 3M then bought and distributed reprints of the articles it wrote and got published and used them to begin increasing doctors'd awareness of their upcoming product launch. While it is unlikely that journal editors were aware of the multiple publication, it is clear that there were not sufficient safeguards in place to prevent it.

This effect was amplified by the way that doctors got their information. Winkle was not smeared in public as a result of publishing his concerns; however, 3M broke off talks with him about conducting future trials and he was never contracted any company to conduct industry research again. In contrast, doctors such as Woosley and Morganroth, who were enthusiastic about the prospects for antiarrhythmic drugs, continued to get funding and, as a result, net numerous publications. In so doing, they became highly influential members of the cardiology community; they were asked to write textbooks, organize symposia, and sit on the FDA advisory panel. Moreover, one of the primary ways that doctors stay up to date is by attending continuing medical educational seminars that describe newly available drugs. When an antiarrhythmic drug gained FDA approval, it was researchers like Morganroth that 3M sponsored to fly around the country and educate doctors about the merits of antiarrhythmic drugs.

[20] As discussed in footnote 5 the loss of advertising is a significant possibility. Marcia Angell (former editor in chief of *NEJM*) explained that pharmaceutical companies do not even have to be involved explicitly; simply the threat of legal action and potential loss of advertising revenue can lead journal editors to self-censor (Abramson, 2004, p. 113).

The value of Morganroth to 3M can hardly be overstated. Due to safety concerns, the FDA approval was for cases of *symptomatic* arrhythmia. While pharmaceutical companies could not legally advertise their drugs for the larger market of asymptomatic cases, they could finance doctors such as Morganroth who recommended such practices. While Winkle did participate in some continuing medical education presentations, it was nowhere near the nationwide junket that manufacturers organized for experts extolling commercially beneficial positions. Again, there is no reason to believe that Morganroth was anything less than sincere in his advocacy, but because his view suited the interests of 3M, it was voices like his that 3M amplified with both speaking engagements and funding for further research. Unfavorable views such as Winkle's existed, but without the amplification of corporate sponsorship. With such an imbalance in the confluence of interests, critics were not so much silenced as drowned out.

This disparity was created by more than just continuing medical education. The marketing department also planned numerous company-sponsored symposia to ensure that journals had a steady stream of articles about their product. These were results that focused on VEB suppression and patient drug tolerance. The results of 28 trials that had attempted and failed to provide evidence for the arrhythmia suppression hypothesis were not widely broadcast, nor did manufacturers trumpet company trials that had been discontinued due to higher death rates.[21] If the NIH had not convened the CAST studies that demonstrated antiarrhythmic drugs increased mortality rates, it is unclear how long it would have taken before such research was conducted. After gaining FDA approval, while companies such as Bristol-Myers Squibb were conducting clinical studies, the vast majority were not addressing pressing unanswered research questions. Instead, most studies were designed to "allow selected

[21] For example, the IMPACT study was stopped due to excess death in the treatment group, but before the results became significant. When the issue of safety arose at the FDA advisory panel, Woosley dismissed the concerns raised by the clinical trial expert from the NIH, saying: "I have a great deal of difficulty taking nonsignificant changes in nonsignificant studies and trying to make significant conclusions" (Moore, 1995, p. 129). This does not take into account the fact that the IMPACT study was designed to be discontinued before such a standard could be achieved.

cardiologists to obtain premarketing experience" and the ability to "compare the effects of Enkaid to previous antiarrhythmic therapies" (as cited in Moore, 1995, p. 179). Bristol-Myer recruited 191 doctors, each of whom was paid a few thousand dollars for their participation.[22] While companies spent between $50 million and $80 million on pre-approval clinical trials,

> a large fraction of this [was] spent on the trials of patients with asymptomatic [VEBs] described above and on large "seeding trials" to place the drugs into the hands of selected "opinion leaders" and "high prescribers" throughout the United States. (Woosley, 1990, p. 556)[23]

Without a control group, such studies could not possibly have told the company anything of scientific value. Of course, that is not to say that such data are of no value *simpliciter*: as noted by CenterWatch, these trials "offer sponsors the opportunity to initiate and develop strategic relationships, especially with high-volume prescribers" (2002, p. 1).

From development to product launch, we can see how commercial forces were instrumental in shaping doctors' perceptions about VEBs, subverting regulatory authority, and amplifying the impact of commercially favorable views. The problems are complex, widespread, and deeply entrenched. Once more, there is nothing particularly unusual about this case; it merely illustrates the pervasive influence of industry funding on the constitution of medical knowledge. Similar factors have been shown to influence prescription habits in a wide variety of cases (Abramson, 2004; Bass, 2008; Brody, 2007; Elliott, 2010; Healy, 2012; Kassirer, 2005; Moynihan & Cassels, 2006). On one hand, no single remedy can address these problems; on the other, an account that does not even recognize commercial influences on the constitution of medical knowledge cannot begin to gain any purchase on them. With a fuller

[22] Precise numbers are not available, but this figure is consistent with industry averages for such arrangements.

[23] Woosley wrote this in the sobering year after the CAST I trial was published in 1989. It should be noted that pre-approval seeding trials were specifically prohibited by the Kefauver-Harris amendments in 1962. Though Woosley was employed by pharmaceutical manufacturers he was also on the FDA advisory panel; thus on this basis alone he had reason to vote against approving antiarrhythmic drugs, or in some way censuring the manufacturers in 1984.

account of antiarrhythmic drugs on the table, I will argue that the philosopher's focus on mechanistic reasoning fails to identify the main threats to epistemically sound treatment decisions.

2.4 The wrong argument: Why a friction-free account of causation does not help in the real world

While no system is infallible, the case of antiarrhythmic drugs exemplifies some of the main problems that commercial forces present for obtaining reliable knowledge. As these factors are ubiquitous, an account of evidence that fails here will similarly fail to deal with persistent threats to best practice. In support of this conclusion, I will argue that if the source of the evidence is neglected, antiarrhythmic therapy was warranted prior to the completion of the CAST I study. If you are of the opinion that information cannot be evaluated independently of the source's reliability, then you are already halfway to my position. In the next section, I will address the possibility of considering commercial bias as one amongst many sources of bias. However, first I will conduct an assessment of the prospects of friction-free epistemology.

According to Cartwright (2007, 2009, 2010, 2011), warrant has three requirements. The first is reason to believe that the effect observed in an RCT is the product of a capacity that will endure outside the experimental setting. For our purposes, the effect observed in the RCT is the ability of antiarrhythmic drugs to prevent VEBs. This effect was well attested to and was the basis on which the drugs were approved. Moreover, a diligent cardiologist could confirm the effect with a Holter heart monitor.[24]

The second requirement is that there is reason to believe the proper causal structure has been identified. On the one hand, we know with the benefit of hindsight that the proper structure was not identified and that preventing VEBs would not yield a decrease of cardiac arrests. On the other hand,

[24] The Holter heart monitor had become a popular way to diagnose VEBs and could also be used to ensure the drugs suppressed VEBs.

several epidemiological studies had shown that VEBs are associated with a higher risk of cardiac arrest. Moreover, the medical journals available to the doctors at the time were flush with articles endorsing the theory. Any cardiologist that attended a continuing medical education event on VEBs would have learned both when to use antiarrhythmic drugs and the supporting causal theory from renowned experts (e.g., Morganroth, Bigger, etc.), and this gave the general practitioner and even the general cardiologist sufficient reason to believe that the correct causal structure had been identified.

The last requirement—that the theoretical description specifies how the cause will contribute to the projected consequence—is intended to identify patients for whom the general causal account will not hold; viz., to identify the defeaters. As Cartwright emphasized, the causal account provides doctors with sub-groups of patients who should not use the drug. For example, some people lack the enzyme that breaks down the antiarrhythmic Enkaid. The Enkaid molecule still had antiarrhythmic properties, but due to the lack of the enzyme, a patient's dose levels would need to be adjusted to ensure proper plasma levels, or another drug substituted. Similarly, use of antiarrhythmic drugs was contraindicated when patients had second- or third-degree AV blocks, a bifascicular block, or a known hypersensitivity to the drug (Anderson, Stewart, & Crevey 1984). Aside from these exceptions, doctors could give the standard account of blocking VEBs, explain how blocking VEBs would prevent cardiac arrest, and confirm that the antiarrhythmic drug was having the intended effect on the person.

Here we have all that Cartwright claimed is necessary to warrant usage of what we now know, and could have known then, is a highly toxic and dangerous drug. Moreover, as I have argued above, the reasons that antiarrhythmic drugs came into use are not idiosyncratic, but represent enduring threats to medical epistemology and thus to safe and effective clinical decision-making. While Cartwright provided valuable insight into the role of causal powers, her account did not address the main threats that face medical epistemology. Howick (2012) suggested a friendly amendment to Cartwright's position by distinguishing "high-quality" causal mechanical reasoning, and suggested that

the antiarrhythmic drug disaster would have been avoided had such a standard been in play. In the next section I will show that mechanical reasoning was not primarily responsible for the disaster, and thus Howick drew the wrong historical lesson from his paradigm case.

2.5 The wrong standard: Why a friction-free account of quality does not help in the real world

Howick (2012) gave two conditions for causal mechanical reasoning being "high quality:" the knowledge of the causal chain from the intervention to a patient-relevant outcome must not be incomplete, and the causal story must take into account the complex nature of human physiology. In Howick's view, doctors were not warranted in prescribing antiarrhythmic drugs because evidence of the causal link between suppressing VEBs and reduced mortality was not high quality. Again, recall that according to Howick's account, if doctors had assessed the effect on mortality with an RCT rather than relying on mechanistic reasoning, the tragedy described above would have been averted. I shall consider two responses to this. The first is to dispute the claim that the available evidence was not high quality. The second is to dispute whether flawed mechanical reasoning is the primary threat to epistemically sound treatment decisions.

Howick (2012) suggested that the reason why mechanistic reasoning was not high quality was that the mechanism that led from VEBs to death was not well supported (p. 127). The only justification he provided was repetition of Moore's (1995) incorrect assertion that the justification for the arrhythmia hypothesis was based solely on large-scale epidemiological studies (see footnote 16). It is supposed that these studies left open the possibility that VEBs were merely a sign of cardiological damage, not a cause of ultimate cardiac arrest. Again, we now know that antiarrhythmic drugs do not decrease mortality, but current knowledge cannot be considered evidence that doctors 30 years ago were unfounded in their beliefs.

To begin with, Howick's suggestion that cardiological damage could be a common cause of VEBs and mortality was examined and researchers established that the relation between VEBs and mortality

held even after controlling for the structural integrity of the heart (Bigger et al., 1984; Lown, 1979; Mukharji et al., 1982). Further, the invention of new monitoring devices had provided unparalleled insights into cardiac arrests. As Holter monitors proliferated, some patients happened to be wearing one at the time of death and this allowed doctors to examine the record of the patients' heart activity just before they died and establish the mechanism that led from VEBs to death. The monitors showed that a VEB occasionally precipitated electric instability and caused the heart to begin beating rapidly and erratically (ventricular tachyarrhythmia), eventuating in electrical cardiac death. Through an examination of the records of patients that died while wearing a Holter monitor, doctors established "that *the mechanism of this condition* in approximately 80% of patients is an acute ventricular tachyarrhythmia that leads to fatal ventricular fibrillation" (Morganroth, 1984, p. 673, emphasis added).

This would seem to be high-quality evidence that VEBs can precipitate death by causing the heart to become electrically unstable. The last key to the mechanistic causal story is to show that antiarrhythmic drugs could prevent ventricular tachyarrhythmia, which is the causal intermediary between VEBs and death. Under controlled conditions, ventricular tachyarrhythmia could be electrically induced by programmed ventricular stimulation. In many cases, the capacity of direct stimulation to evoke tachyarrhythmia was blocked by antiarrhythmic drugs (Anderson, Lutz, & Allison, 1983). Thus, in contrast to Howick's suggestion, there was high-quality evidence regarding the mechanism linking VEBs and cardiac arrests and the ability of antiarrhythmic drugs to prevent cardiac arrest. It is certainly true that the prevention of cardiac arrests had not been shown via RCTs, but if this is the standard of high-quality mechanistic evidence then mechanistic reasoning is not a separate source of evidence; it is just the demonstration of an effect via RCTs.

More importantly, the discussion in Section 2.3 shows that the reason that doctors were prescribing antiarrhythmic drugs was not a reliance on faulty mechanical reasoning. Further evidence that doctors did not rely on mechanical reasoning is provided by the medical community's reaction to

the CAST trial conducted by the NIH. If doctors were relying on mechanical reasoning then, as the mechanisms of class 1 antiarrhythmic drugs were roughly equivalent, the CAST results should have dramatically reduced the prescription of all class 1 antiarrhythmic drugs. Indeed, the CAST I study specifically warned that since the pharmacological properties of all class 1 antiarrhythmic drugs were similar, doctors should not take their patients off of drugs that had been included in the study (Enkaid and Tambocor) simply to put them on a related drug (CAST, 1989; cf. Hine et al., 1989).

In contrast, manufacturers of drugs such as Mexitil that were not included in CAST scrambled to claim the market share lost by Enkaid and Tambocor. They sent thousands of doctors a letter making them aware of the CAST results and sent them a dosing schedule for their convenience if they wished to switch their patients over to Mexitil. Similarly, Parke-Davis' promotional campaign proclaimed "There is no *conclusive* evidence to indict all 1-C arrhythmics" (emphasis in original). Thus, if mechanistic reasoning is driving prescription habits then every drug should decrease in sales, while if marketing is driving these habits then we might expect a much smaller fallout for drugs not included in the study. In fact, while the sales of Enkaid and Tambocor did decrease, sales of other arrhythmics did not. For example, sales of Mexitil *rose* 45% in the 12 months after the CAST I study was published (Moore, 1995, p. 239).

Similarly, CAST II explicitly concluded that the arrhythmia suppression hypothesis was incorrect (CAST, 1992). If doctors were prescribing on the basis of the arrhythmia suppression hypothesis, this would have been the end of the line. The number of deaths would have been astronomically high but it would have stopped growing, and it was the failure of achieving the latter that Morganroth, the erstwhile champion of treating VEBs, found so troubling: "Pick any number [of deaths] you want—50,000, 500,000 … The issue is, which is much more important, is that despite all these articles …

physicians have ignored it [the results of CAST I and II] because sales of quinidine [the oldest and most dangerous arrhythmic] stayed where it was or went up" (Morganroth, as cited in Moore, 1995, p. 246).[25]

The failure of the CAST study shows the power of marketing to overwhelm sound evidence. Clearly, the mechanistic reasoning was flawed, and it is obvious that patients were treated improperly. However, it is just as clear that patients were not treated improperly *because* of flawed mechanistic reasoning, as flawed treatment persisted after the causal story was abandoned. Given that one of the goals of the philosophy of EBM is to identify threats to forming reliable and accurate beliefs from the available data, a friction-free epistemology that ignores pervasive commercial forces will inevitably misidentify the causes of medical folly and fail to provide sound recommendations for choosing treatments.

2.6 Unacceptable answers: Commercial forces are not just one of many confounders

While in most cases philosophers simply ignore the fact that medical evidence is produced in a context of heavy industrial funding, there were occasions on which Howick (2012) acknowledged the problem. By his account, in addition to evidence that a treatment is the best available option, there must also be evidence that the size of the effect outweighs possible confounders. Industry funding comes into his account as one possible confounder; however, this is inadequate for two reasons: first, it understates the role that commercial forces play in medical research, and second, commercial forces are qualitatively unlike other possible confounders, in that they are dynamic.

Howick's (2012) most explicit treatment of commercial forces came in a small section discussing SSRI antidepressants. He noted that while meta-analyses show SSRIs are superior to a placebo, the use of such drugs is not clearly warranted. This is because the modest effect size is not large enough to

[25] It does no good to object that the CAST study happened 30 years ago. The continued use of statins after the ALLHAT study in 2001 and continued use of antidepressants (especially in children) are modern-day examples of irrational prescription due to heavy promotion.

ensure that it could not be accounted for by "publication bias, funding source bias, and data mining in the original studies" (Howick, 2012, p. 59). This frames conflicts of interests or funding source bias as one of many possible threats (along with selection bias, performance bias, etc.) to estimating the correct effect size. As I have already spent considerable time discussing the multifarious ways in which commercial forces pervade medical research, I will only consider two short examples to reinforce the claim that the "one confounder amongst many" view is unsustainable, and discuss four ways in which commercial forces distort treatment decisions beyond skewing estimates of clinical efficacy.

As a first example, let us assuage Howick's concerns and suppose that the *therapeutic effect* of SSRIs is superior to placebo even after controlling for the effects of publication bias, funding source bias, and any other such bias. First, with the active support of industry, the very conception of depression has been radically expanded in the past 20 years, most recently removing the grief exemption in the fifth edition of the DSM (Horwitz & Woolfolk, 2007) (such that if a patient had just suffered the death of a spouse, for instance, we might resist the diagnosis of depression and administration of SSRIs even if "symptoms of depression" lasted for more than two weeks and the patient met current standards for diagnosis). Second, it is not just therapeutic effect size that is skewed, but adverse side effects as well. For years, industry-funded studies found that the incidence of sexual dysfunction among SSRI users was 2–5%, when the true incidence rate is roughly 60% (Glenmullen, 2000). If this example seems trivial, note that substantial underestimation also occurred with the risk of suicide (Healy, 2004).

While these discrepancies are problematic, the major shortcoming of treating commercial forces as one confounder among many is the focus on a "conflicts of interest" model for the role of industry. The idea is that studies will be actively manipulated to yield commercially acceptable outcomes. This does happen, and it is a significant threat to medical knowledge, but there is a more elusive threat of *imbalanced confluence of interests.*

For the second example, we can go back to the arrhythmia case. It is evident that in the early 1980s there was substantial diversity in the opinions being expressed about the potential role of antiarrhythmic drugs. However, industry was able to direct massive resources to researchers that were already inclined towards commercially beneficial positions in order to increase the chance that such views would be accepted. Charitably (and perhaps realistically), we can attribute genuine conviction to every researcher. In other words, Morganroth was as noble in his intention to prevent cardiac arrests in patients with asymptomatic VEBs as Lown was in his conviction that doing so would have dire consequences. Part of what explains the eventual triumph of Morganroth's view was an imbalance between a multibillion dollar industry promoting and fostering his research program, and the relative paucity of opposing forces supporting research that was intended to bring out the possible harms of such treatment (cf. Avorn, 2004, pp. 307–308).

In his advice to pharmaceutical manufacturers, veteran drug developer Lawrence Friedhoff gave the following rule of thumb: "Don't investigate questions unless you can accept all the realistically possible answers" (2009, p. 97). The problem here is not just that some results become skewed by "funding bias," but that research agendas are dominated in such a way that some questions never get asked. However, knowing the commercially unacceptable answers is often crucial for proper medical care. As long as doctors were willing to prescribe antiarrhythmic drugs on the basis of VEB suppression, no company has any incentive to spend considerable funds to run a long-term trial. From a commercial perspective, a negative result will hurt sales and a positive result will merely support the status quo. Thus, the third problem with treating industry funding as one of many confounders is that results are not just biased, but the very aspects of the world that researchers investigate in the first place are systematically skewed. Even if there were accurate estimates of known benefits and known risks, medical research would still yield a distorted picture because questions that are epistemically crucial, but have a high potential to be commercially detrimental, will often go unasked.

Finally, the fourth and most significant way in which funding source bias is unlike other biases is the dynamic nature of commercial influence. In short, the threats that emerge due to the commercial imperatives of medicine are simply manifestations of an underlying drive. Even if—indeed, especially if—researchers find a way to contain an epistemically undesirable manifestation of commercialism, this does not solve the problem. It merely creates an even greater pressure for manufacturers to find a new way to game the system. This contrast can be made clear by distinguishing between error and manipulation.

2.7 Error vs. manipulation: The case that industry does not simply make mistakes

It is worth reiterating that this chapter has not argued that friction-free epistemology is bankrupt. Such analyses serve to articulate an ideal of an account of the logic of inference, and in Chapter 3 I will engage in just such a project. The claim here has been that such accounts represent idealized epistemic positions and abstract away from forces that are entrenched and pervasive in actual instances of medical research. Moreover, the commercial forces that friction-free epistemology abstracts away from raise problems that are significantly different.

For the sake of clarity, we might separate bias into two kinds: *error* and *manipulation*. Error is static. For example, one of the reasons why hormone replacement therapy appeared to have beneficial effects (e.g., preventing cancer) was that healthier women were more likely to seek out treatment in the first place. This possible source of error can be controlled by strictly enforced random allocation of patients at the inception of a medical trial. In addition to employing methodological constraints to minimize the likelihood of making a mistake, researchers may also employ checks to confirm that they have not made any.

This is the guiding idea behind, for example, Deborah Mayo's theory of evidence, and the RCT is one of her canonical examples of arguing from error (Mayo & Spanos, 2010). The general thrust is that

we should be designing procedures that regularly detect the presence of errors. In cases in which such procedures do not detect errors, this provides evidence that the errors are absent and warrants the initial inference. For example, the control group in an RCT is established to guard against the error of mistaking spontaneous improvement and placebo effects for treatment effects. In order to guard against mistakes, Mayo and Spanos (2010) provided a list of canonical errors that threaten the reliability of inference (e.g., mistakes about causal factors). Once a way to control or check for the error has been identified, the task of the epistemologist is complete. In short, the goal is to identify the perfect method.

Manipulation is dynamic. For example, the FDA was created to keep dangerous drugs off the market and pharmaceutical manufacturers have responded by funding the research of commercially friendly academics to produce the required evidence (e.g., Bigger, Woosley, and Pitt). The FDA has placed strictures on what companies can say to doctors in promoting their products, and pharmaceutical manufacturers have responded by paying researchers such as Morganroth to travel around the country putting forward the view that the companies themselves are prohibited from technically endorsing. The dismay caused by cases like the antiarrhythmic drug disaster, in which doctors did not change their prescription habits even after the evidence showed the product was dangerous, have led the EBM movement to train doctors to trust the results of RCTs. Though manufacturers resisted this move in the past (Carpenter, 2010), industry now strongly supports EBM and the reliance on RCTs, in no small part because 75–90% of all RCTs are currently industry funded (Angell, 2004).

2.8 Conclusion

In this chapter I have argued that commercial forces must be included in any discussion of medical epistemology that intends to apply to actual evidence. In Chapter 3, I articulate a fuller theory of the nature of medical epistemology. The general point is that by frustrating your opponents' objectives, you

increase the pressure for them to find a novel response. There is no incorruptible system, but only the ongoing process of move and countermove. This dynamic is the fourth and most essential reason why industry bias cannot be considered one confounder among many. As will be discussed further in Chapter 4, what was discovered during the first attempt to regulate the pharmaceutical market remains true today; that is, that "the difficulty has been, and always must be, the fundamental antagonism between objectives that are largely commercial on the one hand and purely scientific on the other" (CPC, 1920, p. 1,235). New institutions may curb or eliminate an abuse, but few solutions are permanent, and so the epistemologist's task is ongoing.

An antagonistic conception of medical epistemology allows us to see that the ethical dimension of medicine is not just a question of how best to apply what is known, but also one of how medical knowledge is produced or constituted in the first place. There are ethical consequences to allowing manufacturers to play a dominant role in the composition of medical knowledge, and those consequences (which must be considered) are completely opaque on the dominant, EBM, conception of medical epistemology. Bringing the ethical issue to the fore requires an epistemology that is concerned with more than just controlling for error.

As illustrated by the arrhythmic drug disaster, a reliable medical epistemology cannot simply take evidence at face value. It was not by chance that researchers who favored an expansive treatment indication became prominent figures in the cardiology community. Nor was it purely due to error that researchers in influential positions minimized the likelihood that arrhythmic drugs would be found dangerous. It was also not a mistake that rank-and-file doctors were more likely to encounter research that supported the efficacy and safety of these drugs. An epistemology that abstracts away from the "sociological factors" and considers only ideal RCTs is one that can only address one type of bias. A friction-free epistemology simply does not work in a friction-filled world.

CHAPTER 3

Cuckoo Eggs, Improvised Explosive Devices, and RCTs

Towards a General Theory of the Asymmetric Arms Race

3.0 Preface

In the previous chapter we examined the limits of "friction-free epistemology." This chapter explores an example of this epistemology, illustrating that such work can make important contributions as long as philosophers operate within certain boundaries. As such, Chapter 2 is largely an argument against ignoring commercial drivers of medical research.

In this chapter I begin to lay the groundwork for the positive thesis of the dissertation; that is, that medical epistemology is best conceived of as an asymmetric arms race.

To do so, I will briefly stray from medical epistemology to consider a broad class of strategic situations in which two parties with conflicting and mutually exclusive goals respond to each other in a series of moves and countermoves. At this level of abstraction the content area is less important than the pattern and character of the strategic situation. After solidifying the nature of this type of interaction, I will argue that medical epistemology is an example of it—a claim I will further substantiate in the next chapter.

3.1 Introduction

> *A complex weapon makes the strong stronger, while a simple weapon—so long as there is no answer to it—gives claws to the weak.*
>
> (Orwell, 1945)

Each spring in Wicken Fen, reed warblers begin building nests in anticipation of their annual brood and cuckoos sit from nearby perches waiting to slip one of their own eggs in amongst the warbler's clutch. If the cuckoo's egg is allowed to hatch, it will bring forth a ruthless creature. Whenever the blind nestling bumps into another young bird, it will carefully balance its nestmate on its broad back, move to the edge of the nest, and with a swift push from its legs, eject the chick from the nest. The task is not easy for the blind and naked newborn. If the nestmates have not yet hatched, the eggs will take only a few minutes

of effort—though frequent breaks are still required. If the other eggs hatch first, the squirming nestlings will take several attempts over multiple days to eject the young warblers, but eventually they too will be cast into the water below (Dunn, 1985). Accordingly, reed warblers who can detect such threats enjoy a large gain in relative fitness, as do cuckoos that can avoid being detected. Given that there are many possible distinctions and that mistakes will be made, strong evolutionary pressures influence which strategy reed warblers will employ to reliably raise their own chicks.

Two thousand miles away, a group of American troops board their patrol vehicle, turn on their radio jammers and begin visually scanning the road ahead in search of any sign of a newly emplaced improvised explosive devices (IEDs). The scorching Iraqi desert seems worlds away from the placid marsh, yet Dawkins and Krebs (1979) famously asserted that they have much in common. Both the soldier and the reed warbler are engaged in an ongoing struggle against an evolving threat and have adopted new behaviors as a result, and recently this thesis has seen a sustained treatment (Emlen, 2014). Trained as a biologist, but drawing from a number of texts in the evolution of armed conflict, Emlen made the case that biological arms races share the same structural features as military arms races. While I will defend this claim as well, I will argue that Emlen was remiss in discarding the distinction between asymmetric and symmetric arms races.

First, I will examine a military asymmetric arms race in order to identify the features of an asymmetric arms race in general. In Section 3.3 I will show that similar features plausibly exist in biological realms. In Section 3.4 I will explore Emlen's attempt to eliminate the distinctions drawn by Dawkins and Krebs, and argue that such an elimination is inadequate in both the biological and military cases. In Section 3.5 I will solidify the claim that asymmetric arms races constitute a general class of strategic situations by showing that developments in medical epistemology exhibit the same qualitative features, and can thus best be understood in this light.

3.2 The asymmetric war in Iraq: IEDs and MRAPs

> *Revolutionary warfare … represents an exceptional case not only because, as we suspect, it has its special rules, different from those of the conventional war, but also because most of the rules applicable to one side do not work for the other. In a fight between a fly and a lion, the fly cannot deliver a knockout blow and the lion cannot fly. It is the same war for both camps in terms of space and time, yet there are two distinct warfares—the revolutionary's, and shall we say, the counterrevolutionary's.*
>
> (Galula, 1964, p. xii.)

The fight against insurgents is commonly described as an asymmetric war without identifying what makes it so (e.g., Adamson, 2007). As is clear from the submarine/anti-submarine warfare between the Americans and Germans in WWII, an asymmetric arms race is not necessarily between opponents of disparate resources (Meigs, 1990). What defines an asymmetric arms race is that, regardless of the reason, agents with conflicting and mutually exclusive goals are compelled to use disparate strategies, often in the form of measure/countermeasure. In Section 3.2 I will elucidate the consequences of such asymmetries using the term *strategy* to apply to the entire collection of actions that an agent takes (or does not take) and *measure or countermeasure* to refer to the individual components of a strategy. Specifically, I will show that asymmetric arms races are inherently dynamic and thus the reliability of any given strategy is inextricably tied to the strategy employed by one's opponent. While this may be said of many different dynamic games, asymmetric arms races have the following set of features: (1) the reliability of any strategy (once it is employed) typically decreases over time; this is because both (2) opponent responses often attenuate the efficacy of one's strategy and (3) opponents engage in a search process to identify and exploit weaknesses; however, (4) because measures are costly it is often disadvantageous to adopt new strategies until they are necessitated by an opponent; and (5) the process results in the gradual accumulation of costly measures. These five features will serve to ground

the argument in Section 3.3 that an asymmetric arms race is a type of phenomena that spans the natural world and can serve as a canonical model for interspecific asymmetric arms races.

3.2.1 The end of major combat operations, the beginning of asymmetric conflict

On May 1, 2003, six weeks after the invasion of Iraq began, President Bush announced an end to major combat operations. In the early days after the fall of Baghdad, insurgents attacked, but posed little threat to American troops (Adamson, 2007). Frontal assaults on better-trained and superiorly armed troops occurred, but were far more costly to insurgents. However, by late summer, patrolling forces in the "Sunni Triangle" began encountering more opportunistic attacks, such as sniper fire and rocket-propelled grenades. While still within the operational capabilities of American forces, the canvas-doored Humvees driven on most patrols were not equipped for battle (Krepinevich & Wood, 2007).

Though identified as a problem as early as May 2003, the impending scope of the IED emergency to come was not yet apparent. In August of 2003, roadside bombs were terrifying but infrequent; six months of battle had left close to 300 soldiers dead, but only 11 were the result of IEDs (Anderson, Fainaru, & Finer, 2005). During the following year, IEDs proliferated to the point of becoming the primary weapon of the insurgency. In June 2004, the commander of US Central Command, General John Abizaid, wrote a personal letter to Secretary of Defense Donald Rumsfeld calling for "a Manhattan Project-like effort" to deal with the emerging threat. From 2006 to 2011, the Joint IED Defeat Organization spent over $18 billion on counter-IED (C-IED) technology, not including the Pentagon's Mine Resistant Ambush Protected (MRAP) vehicle program, which drew an additional $40 billion from federal coffers (GAO, 2012). To put this in context, the total cost (adjusted for inflation) of the actual Manhattan Project was approximately $20 billion (Schwartz, 1998).

In both military and academic circles, debates have raged about whether C-IED efforts were transformative and necessary (Carter & Gilmore, 2012; Gayl, 2008; Lamb, Schmidt, & Fitzsimmons, 2009) or massive boondoggles (Cary & Youssef, 2011; Krepinevich & Wood 2007; Rohlfs & Sullivan, 2011). Whereas the goal for the actual Manhattan Project was technological and the criteria for success definite, an evolving enemy makes judging C-IED efforts extremely complicated. Whether or not C-IED efforts are judged to be reliable depends on how one evaluates statements such as: IED casualties remain about the same (circa 2007) in spite of a four-fold increase in IED use in Iraq. On one hand, the effectiveness of a single IED decreased steadily over the four-year period. On the other, massive expenditures failed to reduce the probability that a soldier would die from an IED. Yet, the accumulation of costly measures without a substantial decline in casualties is not obviously indicative of failure. Both the increased number and complexity of IEDs, as well as the accumulation of costly C-IED technology on American vehicles, may easily be seen as a standard consequence of participation in the asymmetric arms race.

3.2.2 Make yourself hard to kill

The Army Corps of Engineers divides the conceptual terrain into the five categories: predict, prevent, detect, neutralize, and mitigate. While eventually all facets would be actioned, early efforts were directed at blast mitigation. In this section I will examine the evolution of armored vehicles from the "light-skinned" (i.e., unarmored) Humvee to the fourth-generation MRAP. The MRAP is an important case study because unlike some technology, MRAPs were available from the beginning of the war. Some warplanners advocated skipping the second- and third-generation of development and moving straight to fourth-generation MRAPs, while others resisted such escalation. In Section 3.2.3 I will show that the decision to upgrade vehicles only occurred when the effectiveness of the current generation had been degraded by new insurgent strategies (Criteria 1, 2, and 3). However, I will first set the stage

by reviewing the increasing cost of American countermeasures (Criterion 5), and in 3.2.2.2 I will examine how the refusal to skip second- and third-generation vehicles illustrated a reluctance to get too far ahead of the enemy (Criterion 4).

3.2.2.1 From hillbilly armor to MRAPS: Accumulating costly measures (Criterion 5)

One of the consequences of not preparing for a post-Saddam Iraq was a lack of the right equipment for US Soldiers. This failure is clearest in the primary mode of transportation—the Humvee. Designed to be an agile form of transport behind front lines, the Humvee was ill-equipped to fight an insurgency without any clear distinction between military zones (Krepinevich & Wood, 2007). The military's unofficial policy was to use first-generation "field expedient" armor, a measure described bluntly by Americanserviceman Thomas Wilson to Secretary Rumsfeld on the latter's tour of Iraq: "Our vehicles are not armored. We're digging pieces of rusted scrap metal and compromised ballistic glass that's already been shot up, dropped, busted, picking the best out of this scrap to put on our vehicles to take into combat" (Rumsfeld, 2004).

Though the problem of IEDs was identified in May of 2003, it was not until November that the threat became serious enough to justify the $14,000 cost of armor kits designed to be added to the Humvees already in the field (second-generation vehicles). Production of armor kits went from 35/month at the end of 2003 to 600/month by mid-2004 (Lamb, Schmidt, & Fitzsimmons, 2009).[26] In October 2004, the army purchased 498 additional third-generation armored vehicles and began turning every Humvee into a second-generation up-armored Humvee (Gayl, 2008).

[26] These add-on kits were largely the "Frag-5" add-ons. Except for some niche requirements, "Frag-6" kits were deemed to add unnecessary weight (Erwin, 2007). Armor kits had themselves gone through a gradual evolution. First-generation (Dec 2003) armor kits included L-shaped doors and flank armor; second-generation (Mar 2004) add-ons included armored wheel-wells and underbodies; and third-generation (Nov 2005) integrated add-ons such as the Marine Armor Kit and the Frag 5 (USMC Vehicle Hardening Information Brief, 2005).

Following an inspector general's trip to the theatre in early 2005, the US decided to expedite "bolt-on armor" kits for existing Humvees (second-generation vehicles) in the short term and, as soon as possible, make a full transition to the third-generation M1114 at a cost of $200,000/vehicle (Gayl, 2008; Lamb, Schmidt, & Fitzsimmons, 2009). Unlike the Humvee, the M1114 was designed to be armored from the ground up and was thus far more durable. The second-generation up-armored Humvee added 750–1,000 lbs. of extra armor to the original vehicle, and consequently reduced the Humvee's payload and mobility, and reduced the operational life of the vehicle from 18 months to a year (Gayl, 2008; Krepinevich & Wood, 2007). The third-generation M1114 restored many of the original Humvee capabilities with significantly greater top, bottom, and side protection compared to the previous generation. By 2007, all of the estimated 21,000 Humvees in Iraq were equipped with some form of upgraded armor. Yet the decision in 2005 to field the M1114 was itself superseded less than two years later when the joint chiefs began the fourth-generation MRAP program. In December of 2006, the procurement of 4,000 MRAPs was approved at a cost of $600,000–1,000,000/vehicle (Lamb, Schmidt, & Fitzsimmons, 2009). In June of 2007, the Pentagon approved a one-for-one replacement of up-armored Humvees for MRAPs (Eisler, Morrison, & Vanden Brook, 2007).

The MRAP was not just armored; it was built to allow soldiers to sustain mine blasts and then return fire. Whereas the low, flat-bottomed Humvee trapped the force of underbelly blaststhe hull of the MRAP was armored, raised, and v-shaped to mitigate and deflect the blast force. The weight and the shock-absorbing rubber seats mitigated upward acceleration that could snap spines and throw lighter vehicles off the road. Fixed and sealed containers prevented cabin items from turning into projectiles. Side armor and ballistic glass was designed to mitigate road-side blasts, and in some models soldiers were seated with their backs along the centerline of the vehicle to maximize situational awareness. Firing ports allowed for engaging the enemy from an armored position. Adding to the cost

was a host of other C-IED features, such as frequency jammers to prevent radio-detonated IEDs (Sinclair, 1996). It was, in short, an embodiment of Gen. James Mattis' (former commander of US Central Command) maxim: "Make yourself hard to kill."

While in the beginning of the IED battle, nearly every explosion caused an injury or death, defense officials now stressed the increased protection of fourth-generation vehicles. When sustaining a blast, "nine times out of ten there are no injuries in an MRAP other than bumps, bruises and scrapes. And we're talking about sizable amounts of explosives" (Osborn, 2008, p. 11). Thus, through every generation of vehicle, the military assumed greater costs as new vehicles carried over many developments from earlier generations and made further costly additions.

3.2.2.2 For lack of a crystal ball: Avoiding "unnecessary" costs (Criterion 4)

By designing a vehicle to protect against IEDs from the ground up, military engineers were able to mitigate or eliminate several of the unique threats of IEDs. However, MRAPs were not built to respond to the IED attacks occurring in Iraq; they had been developed in the 1970s in response to the IED threat during the Rhodesian bush wars. MRAPs had been available to American forces from the beginning and the failure to field them when the threat of IEDs became clear requires an explanation.

The argument for each generation of vehicle is plain: each upgrade is an improvement over the earlier generation that it replaced. The more complicated question is why the military failed to field the MRAP earlier. In February of 2005, a universal needs statement was issued by Brigadier General Dennis Hejlik that asked "to utilize supplemental funding to replace 1st/2nd generation vehicles, by skipping a generation and procure 4th generation MRAP vehicles" (as quoted in Gayl, 2008, p. 13) In no uncertain terms, the commander of troops of Multi-National Force West (MNF-W) requested MRAPs, yet the Pentagon ultimately decided to expedite bolt-on armor (second-generation) and field the third-generation M1114.

While a justification for expediting field armor might have been justifiable on the basis of speed, the reasoning behind forgoing MRAPs was far more complicated. The major drawback of the MRAP was the price tag. While Gayl (2008) noted that some of this cost would have been defrayed by the greater life expectancy of the MRAP, there were several other areas that made the MRAP more expensive.

Because of the increased weight, only a few ships and aircraft were able to transport MRAPs, placing increased burdens logistic infrastructure (Erwin, 2007). The heavy weight of the vehicles also resulted in increases in fuel costs (and associated infrastructure to refuel). All told, the three-year operating costs (including transportation, maintenance, and fuel) for the MRAP ($780,000) was over double the cost of the M1114 ($345,000) (Rohlfs & Sullivan, 2011). At the time, side-blast IEDs were the predominant threat to soldiers and the military vehicle hardening study showed that the less expensive M1114 was nearly as reliable in dealing with this threat as the MRAP (Gayl, 2008). As a consequence, the added costs of the MRAP seemed a needless expense. As long as side-blasts remained the predominant threat, it was disadvantageous to field the MRAP because the extra costs were unjustified.

3.2.3 What works today: Measures and countermeasures (Criteria 1, 2, and 3)

The main weakness of the M1114 was the hull. Though the hull was armored, it was flat and low to the ground. It is a simple matter of physics that raising the hull three extra feet off the ground (as in the MRAP) exponentially reduces the force from an underbelly blast (aka deeply buried IED (DBIED)). Predominantly used by Sunni insurgents, DBIEDs had been used before the mass arrival of M1114, but there were only 10 such blasts recorded between January and September 2005 (Gayl, 2008). By August 2006, there were an average of four DBIEDs *per day* in MNF-W alone and by the summer of 2007, DBIEDs were killing more soldiers than all other variants of IEDs combined (Atkinson, 2007). Marine

Corps Commandant James Conway reported that the shift in tactics occurred "because there were more armored Humvees in Anbar" (as cited in Eisler, Morrison, & Vanden Brook, 2007).

DBIEDs were not the only reaction to the M1114: the Shi'a adapted to increased armor by employing explosively formed penetrators (EFPs). An EFP consists of a combination of a bomb and a metal disc that becomes a molten copper slug upon detonation, yet the creation of EFPs requires precise machine tools and adds significant costs to an IED (Wilson, 2006). Even if a vehicle is heavily armored, EFPs cut through the side of it and turned the vehicle armor into shrapnel. Though they emerged in March of 2004, EFP attacks remained relatively rare when far simpler and cheaper IEDs were effective. As second- and third-generation vehicles began to proliferate the battlefield between the spring and summer of 2005, the use of EFPs more than tripled (Rohlfs & Sullivan, 2011). Though M1114s reliably mitigated the predominant threat facing soldiers at the time they were fielded, that very success led to a proliferation of DBIEDs and EFPs. By 2007, DBIEDs and EFPs accounted for 70% of US casualties (Atkinson, 2007), and the inability of M1114s to deal with these devices prompted the Pentagon to approve the complete one-for-one replacement of earlier-generation vehicles with MRAPs.

This shift in tactics is not isolated to the introduction of M1114s. The IED is not just another weapon: "It is not a matter of IEDs on the battlefield, the IED is the battlefield" (Barbero, 2012). IEDs allow small groups with minimal resources to engage in strategic competition effectively and consistently with far more powerful entities. The flexibility of IEDs provided a continued ability to inflict damage against dramatically better-equipped troops, despite numerous attempts to nullify the threat. Further, the ability to detonate devices without direct confrontation decreased the risk to the insurgent and provided the opportunity to learn from failed attempts. All of these factors help to explain why IEDs accounted for over half of the casualties sustained by American troops in Iraq (Krepinevich & Wood, 2007).

As noted in Section 3.2.2.1, American vehicles have gotten progressively more sophisticated; however, soldiers have not become progressively less likely to be killed or injured by an IED. This is because of the corresponding evolution of countermeasures taken by insurgents. This dynamic was exquisitely captured by a US commanding officer: "What works today will not work tomorrow simply because it worked today" (Kirin, as cited in Adamson, 2007, p. 52). First- and second-generation vehicles provided an initial increase in blast mitigation against "double bangers," but increased armor quickly led to "five-bangers" capable of nullifying the added protection of up-armored Humvees (Chisholm, 2005). Just as the Americans did not want to field more expensive vehicles than needed, prior to the arrival of the M1114 DBIEDs were uncommon as they required far more time to emplace, exposing the insurgent to greater risk of being detected by American forces (Criterion 4). Yet increased vehicle armor required insurgents to take additional risks in order to reliably kill American soldiers. Though using DBIEDs led to a quicker depletion of caches of explosives and put the insurgent at greater risk of being detected, as the insurgent began paying the higher cost associated with DBIEDs (Criterion 2) the reliability of M1114s dropped precipitously (Criterion 1). This chapter explores just one aspect of the IED arms race; however, similar adaptations also occurred in trigger mechanisms, detection technology, surveillance, and intelligence.

Upon observing that costly innovations are quickly met with countermeasures that attenuate the reliability of US action (Criteria 1 and 2), a tempting option is to avoid the process of escalation in the first place. Indeed, there was a strong contingent in Rumsfeld's Pentagon that repeated the mantra: "We can't armor our way out of this." Whether the MRAP was worth the investment cannot be determined without access to classified data, but there have certainly been wasteful projects. For example, the predetinator was a multimillion-dollar venture that successfully caused a large majority of

IEDs to explode safely ahead of the vehicle. While extremely reliable in the short-run, once insurgents identified the technology, a 10-cent countermeasure completely negated its effect (Zorpette, 2008).

Projects like these should surely be avoided, yet it is also worth noting that arms races are not just reactive, they are proactive (Criterion 3). It is not as if one side can stop the arms race by ceasing to innovate. Both the military and insurgents are constantly probing each other for ways to better exploit weaknesses. For example, insurgents studied American tactics and began placing IEDs in positions soldiers were likely to occupy after taking small-arms fire (Barbero, 2012; cf. Meigs, 1990). Criterion 3 is a reminder that in such races a failure to continue innovating is a choice to fall increasingly behind.

In summary, the reliability of various forms of vehicle protection decreased over time. This occurred primarily because insurgents began building more expensive and complicated IEDs, but also because they refined their tactics to exploit American vulnerabilities. However, just as insurgents did not assume the costs of DBIEDs or EFPs while cheaper options were largely equivalent, American warplanners refused to field the more expensive MRAP until it was required by the nullification of cheaper options. For both sides, the dynamic has led to the gradual accumulation of costly measures with little change in the number of soldiers killed. As we will see in the next section, such a dynamic is not limited to the battle field.

3.3 What to expect when you're expecting (and you're a reed warbler)

Once a pair of reed warblers has successfully incubated a cuckoo, they surrender the remainder of the breeding season to raise the creature that has killed their young. As a result, there is intense pressure to prevent such a disaster. Though such obligate brood-parasitism (i.e., cases where the parasite species never raises their own young) is relatively rare in the animal kingdom, underlying the interaction is a deeper structure that is far more common. The interaction between the cuckoo and the reed warbler is

a continuously escalating contest between two species in which the success of one comes at the expense of the other. It is a co-evolutionary process that is typified by measures and countermeasures as each species evolve methods targeted to exploit or frustrate the other species' current strategy. As it is commonly accepted that the cuckoo–warbler dynamic is an example of a biological interspecific asymmetric arms race (e.g., Broom, Ruxton, & Kilner, 2008; Servedio & Hauber, 2006; Takasu, 1998), I will confine my discussion to briefly illustrating that the five features identified in military arms races occur in the biological world as well.

The cycle here begins in mid-May, when cuckoos migrate from north Africa to western Europe in order to breed. Instead of building their own nests, cuckoos must find nests of suitable hosts. Upon locating a number of such nests, the cuckoo must keep a close eye on the behavior of the inhabitants. If the cuckoo attempts to lay an egg before the reed warbler herself begins laying, it is virtually certain to be rejected. Similarly, an egg laid after the clutch is completed may receive insufficient incubation and fail to hatch at all (Davies & Brooke, 1988). While only laying during this three- to five-day window dramatically increases the amount of surveillance required, given the reed warbler's strategy, the cuckoo has no choice but to comply. Indeed, cuckoos that miss the window of incubation may depredate a host's clutch to initiate a fresh laying cycle (Gehringer, 1979, as cited in Davies, 2000).

Once a suitable host has been located, the cuckoo waits in a concealed location, such as a nearby tree. When the reed warbler leaves to collect food, the cuckoo descends and alights on the edge of the nest. Though egg laying in most birds takes between 20 and 60 minutes, the cuckoo both snatches a host's egg and lays an egg of its own to replace it in less than 10 seconds. This deft egg exchange is well motivated. Not only can a reed warbler occasionally kill a cuckoo if they spot it, but a cuckoo sighting also causes the warbler to inspect its eggs more closely, and increases the probability of rejecting the imposter egg (Davies & Brooke, 1988).

Yet it is the shells that are the *pièce de résistance* of the cuckoo. Though the focus here is on the dynamic with the reed warbler, there are five other host species. Each host species lays a distinct-looking egg and with one exception (the dunnock), each host will reject non-mimetic eggs (Brooke & Davies, 1988). Amazingly, cuckoos lay different-looking mimetic eggs. It is only for the non-discriminating dunnock that the cuckoo's egg is noticeably distinct from the egg of the host. In the five cases where the host species will eject non-mimetic eggs, the cuckoo's eggs are nearly indistinguishable forgeries. If cuckoos laid indiscriminately, most of their eggs would be detected. Instead, this pressure has led to the emergence of six different gentes of cuckoos. The female of each gens lays one kind of egg and prefers to lay this in the corresponding host's nest, while males mate indiscriminately and keep the population united as a single species (Davies, 2000).

Beyond the shell patterns, cuckoo eggs are especially hard (Brooker & Brooker, 1991) and surprisingly small (Payne, 1974). The cuckoo that parasitizes the reed warbler lays a 3.4 g egg, compared to the 10 g egg laid by non-parasitic cuckoos of the same size (Davies, 2000). The thick shell likely serves multiple purposes. It prevents breakage during the frantic laying process and makes it harder for some hosts to destroy the egg by pecking. It may also increase the cost of ejection by being more likely to cause a breakage in the host eggs, should a reed warbler make such an attempt. The fact that birds without evolutionary exposure to brood parasites will both remain relatively calm in the presence of a stuffed cuckoo and accept non-mimetic eggs in their nest underscores that such behavior is the vestige of prolonged competition between host and parasite (Davies & Brooke, 1989; Moksnes et al., 1995).

Thus, the competition between the reed warbler and the cuckoo is rightfully characterized as an asymmetric arms race. Each population must use disparate measures to maintain a relative fitness advantage over conspecifics. If we suppose that naïve populations of reed warblers, like contemporary

unparasitized populations, first showed no ejection behavior and cuckoos did not lay mimetic eggs, then we can hypothesize a plausible trajectory of the evolutionary dynamic.

First, reed warblers became adept at identifying and ejecting non-mimetic eggs, keeping in mind that because reed warblers lose about one sixth of their clutch to breakage during ejection, this action is costly.[27] Yet reed warblers who reject eggs with non-mimetic shell patterns eliminate roughly 70% of cuckoo eggs. As an illustration of Criterion 1 (decreasing reliability of strategies) and 2 (the emergence of effective countermeasures), the evolution of mimetic eggs by the cuckoo cuts the effectiveness of rejection on the basis of shell pattern to roughly one quarter of its initial effectiveness.

The cuckoo could become even more effective if it evolves the tendency to depredate reed warbler nests during the incubation period to initiate a new cycle of egg laying. Because it is a response to physical realities of incubation time (instead of a countermeasure), it represents an instantiation of Criterion 3. However, just as unparasitized species have no need to pay the cost of occasional mistakes to maintain rejection behavior, the cuckoo gens that primarily parasitizes the non-discriminating dunnock does not pay the cost of making their eggs less acceptable to secondary hosts (Criterion 4).[28] Finally, though the order is not clear, this process repeats and yields an accumulation of discrimination Criteria (timing, size, color) for host rejection and corresponding mimicry, which are just a few of the accumulated costly measures that this arms races has produced (Criterion 5).

3.4 Animal arsenals: Emlen's theory of arms races

The qualitative similarities between military vehicles and reed warblers should give pause. Is talk of "a biological arms race" merely a helpful metaphor, as suggested by Dawkins and Krebs (1979), or is there

[27] The numbers used in this example are taken from those reported in Davies and Brooke (1988).

[28] One cost of laying an egg that mimics that of the dunnock would be an increased rejection rate when the cuckoo lays its eggs in the nests of less-preferred host species.

a deeper similarity in the structure of the interactions. I put forward that an "arms race" is not a mere metaphor for what is happening biologically, but should be seen as a class of strategic scenarios that can occur in multiple domains.

Perhaps the most famous strategic situation is "the prisoner's dilemma" in which two suspects are segregated into separate rooms and incentivized to fink on their partner. If both suspects remain silent then each receives a light sentence, but each suspect can make their situation better—and make their partner's situation considerably worse—by finking. In fact, no matter what their partner does the suspect is better off finking. Even though the total prison time served is lowest if both suspects remain silent, each individual is worse off. However, the prisoner's dilemma is more than an interrogation technique, it is a template for a strategic situation that is faced any time a set of abstract properties are met. Consequently, it has been used to understand phenomena far outside the police station.

Calling a situation "a prisoner's dilemma" is thus not to use a metaphor, but to class the event as having basic structural properties about the "choice" an agent faces. It is in this sense, rather than the metaphorical one, that Emlen (2014) and I discuss arms races. That being said, the aspiration of this chapter is merely to provide a qualitative description of arms races, leaving precise quantification aside for future work.

Beginning from an analysis of animals with large weapons (e.g., the deer's antlers, the fiddler crab's claw, etc.), Emlen (2014) identifies three necessary and sufficient conditions for the initiation of an arms race. First, there must be competition for a good; if there is more than enough of something to go around then there is no use in fighting over it. Second, the good must be economically defensible. Even if a resource is in demand, it must be clumped in such a way that guarding it is feasible. A sparse resource widely and evenly distributed does not create territories worth guarding/fighting to obtain, unless those resources can be concentrated in a new location (e.g., a burrow)—thus making them

economically defensible. Finally, battles between conspecifics must take the form of a duel. In one-on-one combat weapon quality yields a decisive advantage, whereas in "scrambles" the advantage of a superior weapon is minimized by the chaos of the fighting structure.

These requirements are well illustrated by Emlen's first love: dung beetles. Dung beetles have two predominant forms of provisioning dung for their larva. In some species, the females swarm to a new pile of dung, burrow underneath it, and fill the bottom of the burrow with dung from above. Males then occupy and guard the entrances to these tunnels. In other species, the males carve of a ball of dung and roll it away from the main pile (with a female in tow). In the latter case, the provision is desirable enough to lead to competitions and is defensible, but as the male rolls the ball away, he is confronted by numerous other males simultaneously in a "scramble" to own the ball. In contrast, in burrowing species of dung beetles the female and dung lay at the bottom of a thin tunnel, forcing all new comers to confront the current owner in a duel. The addition of the final constraint to dung beetle interactions not only predicts which species have large horns; when species change from one manner of provision to the other over evolutionary history, it also predicts whether species develop or lose horns (Emlen & Phillips, 2006).[29]

The maritime battles of the ancient Mediterranean provide a case example in the military domain. For over a thousand years (1800–750 BCE), oared galleys served simply to transport soldiers to and from the battlefield. Then, the advent of the bronze-tipped battering ram attached to the bow of the boat at the waterline allowed ships to smash the hull of their rivals and sparked an arms race by turning ships into weapons and crews into the power source that fueled duels on the high sea. Since more oarsmen meant quicker ships and since quicker ships won sea battles, the next 400 years saw numerous innovations in naval engineering. The early penteconter had 25 oarsmen per side. Ships

[29] For an impressive list of biological cases that these criteria organize, including fanged frogs and horned gophers, see Emlen, 2014, 92ff.

were first lengthened until their excessive length resulted in buckling. Next, shipwrights began adding tiers of oarsmen stacked on top of each other until the ships became unstable. Finally, ships became wider, with multiple men positioned on each oar. Over the course of the arms races ships had gone from being powered by 50 to over 500 oarsmen.

3.4.1 Sneaks and cheats: Emlen on asymmetric arms races[30]

In some species of burrowing dung beetle, males are dimorphic in that some, like females, fail to develop long horns. At first blush, one might expect short-horned males to be a vestige on the way to extinction, especially when one considers the plight of a small-horned male dung beetle in a burrowing species. Resident males drive their horns into the wall of the burrow, while challengers use their horns to attempt to dislodge the current resident. Small-horned males lack the leverage to eject most residents, and if they happen to find an unprotected burrow they are easily ejected. While long-horned males that get ejected search out another burrow, short-horned males sometimes adopt an alternative strategy.

After being ejected, a short-horned male "cheats." He does not seek out a new opponent that he might best in a fair fight; instead, he moves a small distance away from the original burrow and begins to dig one of his own. After it is deep enough, he climbs inside and remains motionless for several hours. Then, in a frantic burst of activity, the short-horned male excavates a side-tunnel into the original burrow, scrambles down to impregnate the female, and absconds before the resident male is any the wiser. Emlen treats such behavior not as an arms race in its own right, but as a force that can undermine the selective pressures of the original arms race: "when cheaters begin to get too effective, however, they can end an arms race" (p. 156).

[30] This account is drawn from Emlen (2014), especially Chapter 10.

What would seem to be correct is that by circumventing the dueling conventions, small-horned males reduce the benefit that long-horned males achieve. When small-horned males successfully impregnate females guarded by other males, they are reducing the percentage of offspring sired by the resident (big-horned) males. If females cease to be economically defensible, the evolutionary pressure that selected for large horns will disappear as well. Nevertheless, what Emlen seems to fail to appreciate is that this "cheating" is not just a threat to the stability of the original arms race, but an arms race in and of itself. The fact that this is more than just a side-show to the horn-based arms race is evinced by the adaptive behavior seen in both the small-horned (e.g., quicker mating times) and long-horned (e.g., occasional patrol behavior around the burrow entrance) males that are consequences of the arms race between the two dimorphs.

3.4.2 Classifying competitions: Dawkins and Krebs on arms races

While Emlen's framework fails to comfortably accommodate short-horned males, the original framework constructed by Dawkins and Krebs (1979) breaks arms races apart according to whether they are among members of the same species (i.e., intraspecific or interspecific) and whether they are symmetric or asymmetric. In this framework, the competition between long-horned and short-horned males can easily be seen as an intraspecific asymmetric arms race. Similarly, Emlen's discussion of the arms race between fortification and siege warfare is better conceived along the lines of an asymmetric arms race.

Leaving aside the distinction between intraspecific or interspecific, we might return to the symmetric duels between long-horned dung beetles and the high-seas battle on the ancient Mediterranean. Keeping these canonical symmetric arms races in the foreground, we can revisit the five features identified in Section 3.2 to examine which features are common to all arms races. Any

differences between asymmetric and symmetric conflict are further justification that Emlen's treatment is incomplete.[31]

The most obvious common feature is the escalating costs involved in the arms race (Criterion 5). Horn growth comes at the expense of the growth of "eyes, wings, antennae, genitalia and testes depending on the species … males suffer impaired visual acuity, flying agility, smell, and success in copulation" (p. 112). The increased size of ships also represents an obvious increase in costs. It is important to be big, but because of the costs involved, "like animals, the ship size that performed best was relative—'big' really meant 'bigger than everyone else'" (p. 197). Once you are bigger than everyone else, you reliably win duels and being even bigger means paying a higher cost without any gain. In fact, being bigger than everyone else may not maximize reproductive success. Being big enough to win most duels might be good enough, and being even bigger can result in paying extra costs without compensatory gains. Indeed, in several species that are involved in arms races, the animals with the biggest weapons are not as reproductively successful as animals with slightly smaller weapons (Emlen, 2014, p. 251). Just as in asymmetric races, it is often disadvantageous to adopt new strategies until they are necessitated by an opponent (Criterion 4).

The corollary of the last feature is that the effectiveness of a strategy typically decreases over time (Criterion 1). A boat powered by 100 men is a vastly superior weapon than a boat powered by 50 men, but its success sows the seeds of its own destruction, as opponents try to best each successive apex. As with oarsmen, so too with horn size.

[31] Beyond capturing an obvious fact that arms races occur within or between species, it is not clear that this leads to qualitatively different results. For example, Dawkins and Krebs (1979)suggested that a unique feature of interspecific symmetric arms races is that competing species diverge to pursue different strategies in order to circumvent the arms race. This seems to be similar to the evolution of short-horned "sneaks." Further, dung beetles are not alone in polymorphic outcomes to originally intraspecific symmetric arm races (see Emlen, 2014, p. 152f). Whether there is an important qualitative difference will be left unresolved.

But it is here that the similarities end. In an asymmetric race opponents are constrained to use different strategies from one another. The arms race is typified by a series of measures and countermeasures (Criterion 2) with new means of exploitation occasionally discovered (Criterion 3), which then become subject to the same pattern of measures and countermeasures. In contrast, in symmetric arms races, "each new innovation from one side was instantly copied and then bested by the other" (p. 97). Innovation (Criterion 3) occurs in both, but in place of the "measure–countermeasure" dynamic witnessed in asymmetric arms races, symmetric arms races produce "imitation and escalation."

3.5 Epistemic arms races

As claimed above, an asymmetric arms race is a general strategic situation that can be seen outside of military conflict. With an articulated theory in hand, I will support the claim by identifying the same strategic dynamics in a novel domain. Specifically, I propose that any plausible account of medical knowledge must incorporate the fundamental antagonism between those who seek to make medical practice more responsive to good evidence and those whose primary motivations are instead commercial in character.

All arms races are driven by incompatible aspirations. In the medical context, I propose that the antagonism stems from a conflict between pharmaceutical companies pursuing economic self-interest and reformers promoting maximally efficacious treatments. While such a dynamic can explain a number of large-scale developments in the history of medicine, in this chapter I briefly examine the rise of the modern FDA and the rise of a private industry to conduct medical trials for pharmaceutical companies. In the next, I will further substantiate this claim with a case study from the history of medicine.

3.5.1 A new hope: The modern FDA and the codification of RCTs

Thalidomide, the drug most responsible for the creation of the modern FDA, is also an excellent illustration of the arms race between pharmaceutical companies and medical reformers. Thalidomide was given as an antiemetic. Unfortunately, many nauseous patients were also pregnant women and if taken while pregnant, thalidomide dramatically increases the risk of miscarriage and horrifying birth defects, the most well-known which is a defect where arms and/or legs do not fully develop, called phocomelia, or literally "seal limbs."[32] While it was the birth defects that outraged the public, the largest change to the regulatory process was not to safety, but to the type of trial conducted to assess efficacy.

The "trial" that was conducted to assess the efficacy of thalidomide and to earn FDA approval involved 2.5 million tablets sent out to 1,267 doctors for 20,000 patients, and was designed by the marketing department. The detailers securing doctors' participation were each given an "Objectives of the Job" brochure that unequivocally indicated that the purpose was not to generate knowledge, but rather to

> contact teaching hospitals for the purpose of selling them on Kevadon [thalidomide] … doctors need not report results if they don't want to … *Appeal to the doctor's ego—we think that he is important enough to be selected as one of the first to use Kevadon in that section of the country* (as cited in Mintz, 1965, 150f; italics in original).

The congress subsequently heard testimony that such trials were common:

> Much of what passes as clinical investigation from an accounting and advertising point of view is really an effort to get the drug used in a medical center before general release, to get a physician of some influence to use the drug as part of a clinical trial … This I regard as a form of advertising, because I do not think it is a sincere effort to accomplish a clinical investigation of the drug. (Hearings, 1961, p. 10,429).

[32] See Carpenter (2010) for an exhaustive treatment of the episode.

Though, not made illegal, as a result of the Thalidomide disaster such trials were prohibited prior to approval, and commercialization of a drug trial before obtaining FDA approval became grounds for the rejection of a new drug application (Carpenter, 2010, p. 228).

Congress heard detailed testimony that commercialization went far beyond seeding trials into the heart of medical science. The former head of medical research at Pfizer presented the state of clinical trials matter-of-factly:

> A substantial number of the so-called medical scientific papers that are published on behalf of these drugs are written within the confines of the pharmaceutical houses concerned. Frequently the physician involved merely makes the observations and his data, which are sometimes sketchy and uncritical, are submitted to a medical writer employed by the company. The writer prepares the article which is returned to the physician who makes the overt effort to submit it for publication. The article is frequently sent to one of the journals which looks to the pharmaceutical company for advertising and rarely is publication refused … [For example,] I was assigned the task of writing a paper on a new formulation of a broad-spectrum antibiotic. I was informed that this paper had been accepted for publication and the 100,000-plus reprints were ordered before I finished the writing assignment. The paper, or course, was published exactly on schedule, which incidentally was within a few days of the introduction of the product on the market. (Hearings, p. 10,244f)

Each manufacture had generous expense accounts that were used to maintain "stables" of physicians that could be relied on to write laudatory "clinical studies" about new drugs, or at least agree to be the nominal authors of industry-written papers (Hearings, 1961, p. 10,371f; cf. Mintz, 1965).

In response to these abuses, medical reformers altered the drug approval process to separate clinical evaluation from the business of promotion. Instead of practicing physicians assessing efficacy, the bill required "experts qualified by training and experience" to establish drug efficacy by "well-controlled investigations." With the conceptual authority granted to it by congress, the FDA sought to set a high minimum bar for market entry; "experts" became pharmacologists and "well-controlled observations" became RCTs. The sort of trial that had been conducted for Thalidomide was now completely banned.

3.5.2 The industry strikes back: Rise of the CROs

After the changes at the FDA there were numerous consequences for the drug market. The National Academy of Sciences was tasked with reevaluating any drug that had gained market entry prior to the new standards. The results of the study resulted in a major financial hit to pharmaceutical manufacturers: 7% of drugs were removed from the market, while half of drugs had their promotion curtailed because they were judged to be ineffective for at least one of the diseases they had been promoted for (Hilts, 2003). Similarly, new powers at the FDA resulted in the discontinuation of 10% of new drug applications and hundreds of other applications were methodologically strengthened as the result of early warnings (Carpenter, 2010). Finally, the FDA created several advisory boards that brought academics into service for the FDA, established plans to have their doctors spend one day a week in a university, and several other initiatives that the director of medical review intended to "develop a close liaison with between the Bureau of Medicine and the scientific community" (Sandusk, as cited in Carpenter, 2010, p. 307).

In the short run, the FDA strengthened its connections with the academic community and functionally mandated that pharmaceutical manufacturers contract academics to conduct their clinical trials. In addition to a smaller percentage of drugs making it through the approval process, this process began to take considerably longer. This occurred in part because of the added requirements, but also because university researchers were less beholden to company interests. Academic researchers were less inclined to yield to company pressure for rapid completion of trials, less interested in research geared towards meeting regulatory demands, and more inclined to explore intellectually interesting

projects, even if such projects may lead to commercially unfavorable discoveries (Mirowski & Van Horn, 2005).[33]

Beginning in 1974 with IntraMed, a private industry of CROs emerged that was geared towards meeting the needs of industry.[34] Since CROs were dedicated to serving industry, their trials were tailored to winning regulatory approval as quickly as possible. Whereas an academic researcher may have career demands (e.g., gaining tenure) that would motivate them to complete and publish trial results regardless of outcome, employees of CROs were sensitive to pharmaceutical companies' need for cost containment (Abraham, 2002). Whereas in 1990, 80% of industry research was conducted in the university setting, by 2000 the percentage had dropped to less than 40% (Angell, 2004). Faced with competition from CROs, university researchers began to adopt a partnership model, in which researchers "rarely have full participation in trial design, unimpeded access to trial data, and the right to publish their findings" (Schulman et al., 2002, p. 1,335). Moreover, because of the widespread practice of ghost authorship[35], the line between industry and academia has blurred even further.

What is crucial to note is that because studies conducted by CROs are geared towards regulatory approval they are, to all outward appearances, paragons of methodological rigor. Indeed, comparisons of industry-funded trials with independently funded trials find the former to be methodologically superior (Lexchin et al., 2003; Smith, 2005). However, they bear all the marks of rigorous science in the same way that cuckoo eggs bear all the marks of being those of reed warblers. Industry funding has

[33] For example, Dr. Roger Winkle, a professor at Stanford, was contracted in 1980 to test a drug designed to prevent heart attacks. When he noticed that it seemed to precipitate heart attacks in some patients he conducted a retrospective analysis and found a number of such cases. The publication of his results cost pharmaceutical companies nearly 85% of their market share by limiting the patient population for the drug (Moore, 1995). For a general treatment of the differences in how industry and academic scientists regard ownership of data, see David and Dasgupta (1994).

[34] Other early companies include Quintiles (est. 1982), Parexel (est. 1984), and Covance (est. 1987).

[35] Ghost authorship occurs when an author agrees to be the nominal author on a paper describing trials they had no part in. Healy (2004) found that this practice accounted for over half of articles published for the antidepressant Zoloft, and in less than 5% of cases was the involvement of the CRO disclosed. For a more extended discussion of this practice, see Sismondo (2009).

been shown to correlate with commercially favorable results in arthritis (Rochon et al., 1994), hypertension (Stelfox et al., 1998), Alzheimer's disease (Koepp, 1999), oral contraception (Vandenbroucke et al., 2000), oncology (Knox et al., 2000), antidepressants (Healy & Cattell, 2003), schizophrenia (Montgomery, 2004), psychiatry (in particular) (Mandlekern, 1999), and medicine in general (Lexchin et al., 2003; Lundh et al., 2012). What is most worrying about such research is not that the data are fraudulent—though sometimes they are—but rather that the data gives truthful answers to carefully put and misleading questions; that in myriad and subtle ways, the company has gone "to great pains to construct studies that are likely to turn out in its favor" (Gelenberg, 1999, p. 122).

3.5.3 An epistemic arms race

Though this section has focused on RCTs, I will argue in Chapter 7 that the actual competition is far broader and involves medical education, publication, conference organization, and a number of other ways that influence the beliefs that doctors hold. For example, pharmaceutical companies have discovered that paying doctors a small fee to become a "key opinion leader" and lecture on behalf of the company leads that very same doctor to write far more prescriptions.[36] Nevertheless, examining the role of RCTs provides a window into the epistemological arms race, just as blast mitigation does for IEDs.

The driving force in the medical field is prescription decisions. Whenever a doctor writes a prescription (or chooses not to), two things change: the patient's chances of recovery and the profitability of the pharmaceutical. Medical reformers have attempted to maximize the former, while

[36] According to former drug detailer Mathew Webb, asking a doctor to speak about a product was one of the best ways to change that doctor's prescription habits: "I would say 'Can you speak for me on this product?' Because I knew in the back of my mind if I get him to speak for me he's gonna write more product… other people would start using it, but the big bump would come for the speaker." The company would pay $1,500 for the doctor to give a talk and that would result in the doctor writing an extra $100,000 to $200,000 more in that company's product (Spiegel, 2010).

pharmaceutical manufacturers have amplified the latter.[37] Because maximizing these two outcomes often cannot be done simultaneously, there is fierce competition for the doctor's pen and pad. Reformers have increasingly brought medical trials to bear on medical decision making, first during the 1960s through regulation of market entry and then even more so in the past 20 years by encouraging doctors to base treatment decisions on RCTs (this latter is referred to as the EBM movement).

The codification of RCTs can be seen as a strategy used by medical reformers to eliminate the worst drugs from the marketplace, and EBM can be seen as an attempt to leverage the same tools to promote best practice. However, as Frances Kelsey, the reviewer that prevented thalidomide from entering the US market noted in regards to the changes she had witnessed in four decades at the FDA: "We certainly got rid of some of the horrors of the old days, but things change and new problems pop up" (Kelsey, as cited in Fried, 1998, p. 150). As a strategy for improving practice, RCT reliability has decreased over time (Criterion 1). This is due to both countermeasures (e.g., CROs) developed by pharmaceutical companies (Criterion 2) and novel means of influencing physicians, such as hiring them to speak about the product as "key opinion leaders" (Criterion 3). However, it should be noted that while CROs produce commercially beneficial products (e.g., clinical trials) for pharmaceutical companies, they are extremely expensive, costing pharmaceutical companies almost $8 billion in 2001 alone (Mirowski & Van Horn, 2005). Such expenditure was not one that companies accepted until opponents' strategies (i.e., regulatory changes and the EBM movement) required it (Criterion 4). As the above dollar amount indicates (cf. Carpenter, 2010), the cost of both drug development and regulation has increased substantially over the past 50 years (Criterion 5).

[37] Lest a charitable reader think I am being too cynical in my characterization, the fact that a business school education leads to corporate boardrooms that consistently sacrifice public health to shareholder profit has been demonstrated in a long series of experiments (91 trials in 10 countries; Armstrong, 2001). As was eloquently put by a British pharmaceutical firm: "*What we do for the public good is a byproduct of what we do for our own private good*" (emphasis in original, as cited in Lasagna, 1962, p. 146).

3.6 Conclusion

This chapter has provided a general theory of arms races and specified the characteristics of both symmetric and asymmetric arms races. The fact that many of the aspects are present in any arms race does not detract from their importance in any particular case. Moreover, I have defended the claim that rather than being a mere metaphor, arms races represent a class of strategic phenomena by showing that the same deep structural dynamics occur in areas that are otherwise unrelated. Specifically, I have provided an example of how two of the most significant developments in the structure of medical research (the modern FDA and the rise of CROs) can be understood as part of an epistemic asymmetric arms race. In the next chapter I will show that this is not a recent aberration, and that it captures both large- and small-scale changes in medical research.

One further point regarding the consequences of Criterion 4 might be worth highlighting here, as it will be a theme that runs throughout the next chapter. If the arms race picture is correct, then pharmaceutical companies have no interest in distorting the constitution of knowledge so long as what is known has no impact on what is prescribed (i.e., profits).[38] Provided knowledge remains recondite and removed from practice, companies have no desire to interfere. It is only when knowledge is brought to bear on the market that it becomes of interest. As will be shown in the next chapter, RCT methodology was developed in the 1910s, but CROs did not arise until the 1970s and did not attain their ubiquity until the 2000s.

This timeline is fully consistent with the arms race account when one considers what impacted treatment decisions. Prior to the 1960s, clinical impressions of efficacy (i.e., testimonials) were

[38] By "known," I mean to indicate only the maximal epistemic position that an agent could be in given the available evidence at the time, regardless of how many agents actually occupied such a position. For example, it could be said that it was known that Vioxx increased the risk of heart attacks the entire time that the drug was on the market because that is what the best available data justified one to believe. It could be further argued that to what extent such knowledge was *widely* known is irrelevant to epistemology. In Chapter 7 I will consider a different account of knowledge that is far more applicable to the concerns of medical epistemology.

sufficient for both market entry and convincing doctors to prescribe a drug.[39] It was not until the FDA required RCTs for market entry in the mid-1960s that companies had reason to invest significant capital in figuring out a way to subvert the reliability of RCTs. This pressure was significantly intensified in the 1990s as doctors began to be trained to base treatment decisions on RCTs. Once the results of RCTs became directly influential on profits, companies faced significant financial pressure to run RCTs that would support their product. In short, the FDA and EBM created an environment in which, in the words of the CRO Envision Pharma, "data generated in clinical trials are the most powerful marketing tools available to a pharmaceutical company" (as cited in Healy, 2012, p. 104).

[39] In Chapter 6 I will substantiate this claim further by examining a drug (DES) that numerous RCTs showed to be ineffective, but was neither removed from the market by the FDA nor suffered from decreased sales.

CHAPTER 4

The Fundamental Antagonism

Veritism and Commerce in the Early Twentieth Century

4.0 Preface

The previous chapter moved away from medical epistemology in order to establish a general theory of asymmetric arms races. In such races, competing parties with conflicting and mutually exclusive goals are constrained to use a different set of strategies in order to compete with each other. In the resulting arms race the efficacies of strategies decrease over time as one's opponent develops countermeasures and/or finds new ways to exploit weaknesses. The result is the accumulation of costly measures that accrue as part of the race, in which the best place to be is only one step ahead. Accordingly, the dynamic is dominated by a series of measures and countermeasures as opponents seek to undermine the strategies being employed by their opponent. The argument that "arms race" is a general strategic situation applicable to medical epistemology is strengthened in this chapter by exploring another instance of the arms race between pharmaceutical manufacturers and medical reformers.

4.1 Introduction

> *[Austin Bradford]Hill's work had outlined the basic structure within which clinical trials would subsequently be conducted … The Laplacian vision of the determination of medical therapy on the basis of the calculus of probabilities had finally found its spokesman.*
>
> (Matthews, 1995, p. 130)

In *Quantification and the Quest for Medical Certainty,* Matthews (1995) traces back the intellectual history of Austin Bradford Hill—the head of the United Kingdom's Medical Research Council (MRC) when it conducted the first modern clinical trial. In fact, just prior to the famous clinical trial of streptomycin, Hill had assumed both his position at the MRC and a professorship at the London School of Hygiene from his mentor, Major Greenwood. Greenwood was, in turn, a student of Karl Pearson, who himself was the protégé and intellectual heir of his benefactor Francis Galton. Though the line of direct instruction ends here, Galton and the British biometric school are represented as pursuing, at the turn of the twentieth

century, the same project as Radicke had attempted unsuccessfully in Germany during the late nineteenth century, and that had been dismissed in mid-nineteenth century France when promulgated by Gavarret; namely, the medical application of the probability calculus as propounded by Pierre-Simon de Laplace in the early nineteenth century (and as developed by Quetelet and Poisson).

In fairness to Matthews, his book includes fascinating investigations into disciplinary disputes over which profession has intellectual authority in judgments of therapeutic efficacy and debates about the source of objectivity. Yet as Harry Marks (1997) has pointed out, it remains, along with many other histories of the clinical trial, a collection of "disparate episodes … linked in a transhistorical narrative of antecedents deemed to constitute the history of the present day randomized clinical trial" (p. 6). To replace such a view, Marks' *Progress of Experiment* united the work of interdisciplinary efforts into what he called a history of "therapeutic reformers." Specifically, Marks identified the defining characteristics of reformers as "the shared belief that better knowledge about the effects and the uses of drugs will lead to better therapeutic practice" (p. 3). While Marks' contribution is to be lauded for contextualizing the work statisticians, pharmacologists, chemists and other reformers, it remains conceptually problematic. Like the histories Marks critiqued, the history of the clinical trial is treated as an essentially cumulative and progressive endeavor. Such a history entails that as the clinical trial is refined, successive generations are in increasingly better epistemic positions and have *progressed* closer to the attainment of medical certainty.

What is lurking in the background, but never clearly articulated by Marks, is an explicit role for the entities that reformers reacted against and a realization that therapeutic reform required far more than improving scientific methodology. Rather than a gradual progress of experiment, I suggest that medical reformers are better conceptualized as one party in an asymmetric arms race, and that their history cannot be adequately understood separately from the corresponding developments of the other

party—the pharmaceutical industry. By analogy, tracing back the technological and intellectual history of the MRAP without discussing its larger context, its strategic importance, and the impetus for its creation and refinement would be inscrutable without an equally clear understanding of how IEDs threatened American troops.

All arms races are driven by incompatible aspirations. In the medical context, I propose that the antagonism stems from a conflict between pharmaceutical companies pursuing economic self-interest and reformers promoting maximally efficacious treatments. While such a dynamic can explain a number of large-scale developments in the history of medicine, this chapter focuses solely on the first group of reformers examined by Marks: the Council of Pharmacy and Chemistry (CPC).

The council plays a significant, though underappreciated, role in the history of EBM. It was the first group to be dedicated to regulation of the pharmaceutical industry and, as I will argue below, should be credited, rather than Hill, as the actual originator of the double-blind RCT.[40] The council's vision of clinical pharmacology as the gatekeepers of medical practice was the basis of the regulatory model adopted by the US FDA in the 1960s, and worldwide in the decades that followed. In this regard, previous writers on EBM have given the council short shrift. However, even if they were to be recognized for these contributions, the central lessons of their attempts at reform would remain obscured.

Previous histories have focused on one or two innovative studies conducted by the council as part of a larger sweeping narrative tracing the intellectual development of trial methodology in medical research. In contrast, the present work primarily draws from the collection of reports that the council published between the years 1905 and 1920. The result is a dramatically different picture. First, there is significantly less focus on the role clinical testing. More importantly, the essential relation between the

[40] Many different "councils" are mentioned in passing. I shall reserve "council" *simpliciter* to refer to the Council of Pharmacy and Chemistry.

scientific standards employed by the council and the advertising strategies of pharmaceutical companies comes to the fore. As I will argue, constructive research (i.e., establishing new knowledge or methodology) is only half the story (probably far less). The movement led by the council was not merely an attempt to establish a scientific means of evaluating treatments; it was a reaction against a marketplace that was overrun with hucksters, quacks, and miracle cures.

While it is often constructive work that is built upon and elaborated by future generations (e.g., clinical trials), the success of the council's constructive work depended on ephemeral and contextual responses to contemporaneous methods of drug promotion. These contextual responses have been neglected both historically and in contemporary work on medical epistemology, yet were central to changing the standards of knowledge in the medical community. Further, if we see a physician's prescription as a rough proxy for the state of that doctor's knowledge, then it is possible to gauge the collective knowledge of doctors by looking at nationwide sales data.[41] In cases where conclusive evidence shows a drug to be strictly inferior to other available treatments, it is possible to examine why doctors continued to believe the contrary.

Referred to as "irrational prescription," the continued use of worthless products became a central focus for the council. Though it began with the belief that all the profession lacked was access to reliable information, years of struggle taught it that "the difficulty has been, and always must be, the fundamental antagonism between objectives that are largely commercial on the one hand and purely

[41] I will justify this approach in the next chapter. In short, reference to prescription data seems to be a far superior gauge of "medical knowledge" than the results of the most rigorous experiments available at a given time. If an elite group of researchers has established some therapeutic fact, but this is not actioned by the medical community at large, then the therapeutic fact is not known in the most relevant sense (i.e., it does nothing to improve patient outcomes). Surely there are other factors that affect prescription decisions (e.g., cost, availability, etc.), but this seems like a good (and accessible) first approximation. Further, I will charitably assume that doctors prescribe what they believe is in the best interest of their patient, though I also believe this is generally true. It seems that doctors' prescription habits present an excellent example of applied social epistemology; if there are shortcomings with using doctors' prescriptions as an operationalization of their knowledge, hopefully critiques will lead to conceptual progress in the matter.

scientific on the other" (CPC, 1920, p. 1,235). In this chapter I will make the case for, and elaborate the consequences of, recognizing this fundamental antagonism in medical research.

4.2. Pecksniffian Virtue (1905–1910): The General Incompatibility Between Good Health and High Profits

> The council is bound to conflict with the commercial element where this element conflicts with scientific progress; the manufacturers, on the other hand, must keep an eye to dividends. (Sollmann, 1908)

In the years preceding the council's formation, the use of patent medications had exploded from 28 percent of all drugs produced in 1880 to 72 percent in 1900 (Marks, 1998).[42] In principle, the primary difference between the despised patent-medicine makers and ethical pharmaceutical companies was determined by what types of products a company produced and how the company promoted these items.[43] Ethical manufacturers made known ingredients that were sold in bulk to compounding pharmacists to be used when filling doctors' prescriptions. They did not advertise to the public, but instead garnered sales within the medical community via reputations of integrity and quality. Patent-medicine companies were seen as, and often were, snake-oil salesmen who were duping the public with panaceas (Young, 1961). Their products were not actually patented, but rather contained secret ingredients said to be responsible for their remarkable curative powers.[44]

---

[42] Indeed, physicians had steadily begun to transition to writing prescriptions for patent medications, going from less than 1 percent in 1874 to close to 25 percent by 1902 (Jacobi, 1906).

[43] The term "ethical" is used in a nominal sense to refer to companies who publically avowed the ethical norms of the medical community. It is not an assertion that business practices used were in fact more ethical.

[44] It is worth noting that patent-medicine companies competed with doctors for medical authority in the eyes of the public. Moreover, given the state of medical practice at the time, it is not clear that people were better off visiting physicians. Likewise, some patent medicines were efficacious treatments. For example, Dr. Sappington's Anti-fever Pills (secretly) contained quinine and were thus often effective against fevers, particularly in the south where malaria was endemic. Whereas medical reformers viewed secrecy with distrust, manufacturers saw it as necessary to avoid being undercut by inferior rivals—a fate Sappington indeed faced when he succumbed to pressure to reveal his ingredients (for this case in particular, and the evolution of these categories and the respective types of companies more generally, see Gabriel, 2014)). Nevertheless, effective patent medicines

In the late nineteenth century, a type of manufacturer emerged that was neither decisively a patent nor an ethical manufacturer. These companies marketed to doctors, but created readymade products (rather than ingredients) and promoted their wares with all of the garish promises that epitomized patent medicine and medical quackery. It was with the aim of eliminating this penumbra of ethical practice that a small group of men within the American Medical Association (AMA) dedicated themselves to founding The Council of Pharmacy and Chemistry.

Early on, the council generally believed that doctors had been duped and/or merely lacked information to prescribe rationally. Thus, the council's initial actions were targeted at exposing the fraud and deception of manufacturers with the full expectation that sunshine would be a sufficient disinfectant (Simmons, 1906).[45] The first five years of the council's work can roughly be thought of as the optimistic phase. The council focused their efforts on exposing deceitful advertising practices with the confidence that doctors would cease patronizing companies that had been exposed. Moreover, they expected reputable manufacturers to welcome the fact that they no longer needed to compete with such dishonesty. Ultimately, the council was wrong on both accounts. Nevertheless, an examination of their early efforts remains an excellent illustration of the driving force of the arms race—namely, that the satisfaction of the veritistic aims of the council typically frustrated the commercial aims of the products' manufacturers (and vice versa).

The first task the council set for itself was to evaluate an entire category of drugs: *proprietary medicines* (Sollmann, 1908). This interloper was neither a standard ethical product nor an obviously quackish patent drug. Proprietary drugs were ready-made products advertised only to doctors. Though some proprietary medication, such as aspirin, were valuable drugs produced by reputable companies,

---

almost exclusively worked because they included ingredients known to be effective, and which could be bought much more cheaply from a reputable pharmacist, instead of the exorbitant prices charged for the readymade patent drug (Hopkins, 1907).

[45] Hatcher (1916) reports that it seemed almost self-evident that exposure of fraud would lead to its disappearance.

the majority were no better than the typical patent drug pawned off on the public. To entice compliance, the council had "sticks" such as being able to prevent products from being advertised in *JAMA,* and "carrots" such as deciding which products to include in council publications including *New and Nonofficial Remedies.*

The council's primary goal was to deprive from the ranks of respectability the supposedly ethical drugs that where nothing but quack remedies wrapped in the cloak of pseudoscience (derogatively referred to as "nostrums"). A secondary function of this evaluation was to elevate the behavior of the truly ethical firm. Privately, they believed that only a small number of the over 600 proprietary drugs were worthy of any real distinction (Sollmann, 1908).

The cause for the disparity between the council's opinion and general consensus lay primarily in their commitment to rational therapeutics in place of empirical ones. The latter held that doctors should employ what they found to work in their practice and not rely on a dogmatic therapeutic system. In contrast, the council championed an alternative epistemological foundation to determine normative guidelines: laboratory science. Crucially, rational therapists rejected standard clinical experience as a wholly unreliable source of knowledge.[46] Rational therapeutics required that doctors prescribed a specific substance in order to obtain a specific effect (Davis, 1902).

More than just extolling the virtues of science, the actions of the council were tailored to the specific practices perceived to threaten ethical medicine. While most of the their reports were technical analyses of particular drugs, it also added to some analyses object lessons for the profession in rational therapeutics. The campaign against headache powders, discussed below, is an illustrative example of such a lesson. Further, it captures a number of predominant themes that run through the council's

[46] For example, in response to a doctor basing his arteriosclerosis treatment on the beneficial results he obtained in his practice, the council rejoined: "So unscientific is the empirical method that it is hardly worth taking the space to demonstrate its imperfections" (CPC, 1914a, p. 1,035). For an earlier discussion of the imprecision of clinical experience, see Sollmann (1908) and Marks (1997).

early work: it included concerns about deception, secrecy, and direct appeal to the laity by the manufacturer; advocacy of science; and condemnation of the corrupting influence of commercial imperatives. In short, the campaign demarcated the aims of reformers from the aims of business and exhorted America's doctors to frustrate commercial strategies that were fiscally successful, but that came at the cost of patient well-being.

4.2.1 The Most important medium of advertising: Headache powders and the struggle for doctors' hearts and minds

One of the most important commercial assets of a patent medicine was the veil of secrecy concerning its contents. By keeping their ingredients secret, manufacturers could make grandiose claims for products with ordinary components. In an attempt to remove such lucrative façades, one of the first reports put out by the council was a laboratory analysis revealing that a number of headache powders were essentially nothing but acetanilid, a well-known substance that could be purchased for pennies on the dollar (CPC, 1905a).[47] Given the possibility of addiction, poisoning, and death associated with the use of acetanilid, the drugs served as an example of why empirical therapeutics was problematic (Austin & Larrabee, 1906). The acetanilid mixtures illustrated the dire need for doctors to know not only what they were prescribing, but in what amounts. While the council did not attack a doctor's right to prescribe a drug they believed to be helpful, they enjoined that "no physician, however, has any right, either moral or professional, to prescribe a preparation, concerning the ingredients of which he knows

[47] Acetanilid was one of three (along with antipyrine and phenacetin) coal-tar-derived products discovered in the mid-1880s that helped legitimate the role of pharmaceutical companies introducing products into the market. These three raw products should not be confused with products such as Antikamnia, which contained them as (secret) ingredients. The promotion of the former were far more in line with the ethical standards of the day. For an exploration of how headache powders gave birth to the modern pharmaceutical company and shaped professional disputes about the appropriate roles for doctors and pharmacists, see McTavish (2004). Though Antikamnia became one of the first extended battlegrounds for the council, they were not the first to reveal that Antikamnia contained acetanilid (e.g., Robinson, 1904).

absolutely nothing"(CPC, 1907a; cf. CPC, 1905b). This was one of many ways the council framed its alternative epistemic framework as a doctor's ethical duty.[48]

Of the headache cures, Antikamnia became the paradigm of a proprietary nostrum. Shortly after its contents were exposed (a claim the company denied), the federal government classed acetanilid with opium and cocaine among dangerous ingredients that must be disclosed. It seemed the manufacturer would now either have to admit that Antikamnia contained acetanilid or face federal prosecution. Yet rather than do either, the company replaced acetanilid with a more costly, but related and equally dangerous ingredient (phenacetin). Though this ingredient was also required to be listed, Antikamnia manufacturers used an obscure technical term instead of the popular designation (CPC, 1908a). When secrecy became too risky, the company took refuge in subterfuge. In the council's opinion, the only reasons for secrecy were because a product contained an ingredient such as acetanilid that doctors would only use with caution, or that a product used ingredients that had no expected benefit and was thus something doctors would not use at all (Sollmann, 1908).[49]

The council chose Phenalgin, a second acetanilid-containing remedy, to illustrate fraudulent advertising and exploitation of the laity. Like the manufactures of Antikamnia, the makers of Phenalgin hid the fact that it contained toxic components and falsely claimed that Phenalgin was a unique chemical substance. Ads run in 1905 maintained that Phenalgin was manufactured "under the immediate personal supervision of the original inventor," a remarkable feat for a man who had been

[48] These concerns were most pressing for potentially poisonous substances such as acetanilid or iodids, yet they were by no means confined to such substances. For example, the council decried the duplicity of Hagee's Cordial of Cod-liver Oil. Hagee's had cleverly addressed the unpalatableness of cod-liver oil by removing it from the cordial, but not from the name (CPC, 1906a). For an example of iodids see Puckner and Clark (1908).

[49] Though this critique of secrecy was often implicit in criticisms of proprietary medication, it was plainly stated. For example, the makers of Micajah's Medicated Uterine Wafers claimed that their product could cure serious diseases such as uterine cancer and gonorrhea. After laboratory analysis showed the wafers were composed of an astringent and a weak antiseptic, the council noted that the mixtures of common and well-known ingredients "are foisted on the medical profession with no hint as to their composition and with claims made that are not only false, but would immediately be recognized as absurd, if their actual composition were known" (CPC, 1910a, p. 1,216).

dead for two years (CPC, 1906b). The council felt that such deceitful claims were merely an indication of duplicity that extended beyond the ad copy, pervading every aspect of the company.

By taking a common substance (acetanilid) and combining it with starch and a liberal amount of printer's ink, manufacturers of phenalgin had convinced doctors to prescribe a drug in situations where it would not help, and caused patients to pay exorbitant fees as a result.[50] In addition to advertising, manufacturers could prompt wider initial use by distributing free samples or by selling doctors stock in the company in exchange for their agreement to prescribe the drug. The council attempted to harness the animosity that doctors harbored for patent medications, reminding doctors that many had begun as proprietary drugs:

> When the history of the "patent medicine" business comes to be written impartially and fairly, it will be realized that we, the medical profession, have been in no small degree responsible for its growth. Not a few widely advertised nostrums owe their commercial success solely to the ill-considered use accorded them by physicians, to whom they were first exploited. (CPC, 1912b, p. 666)

By including a pamphlet with each bottle, manufacturers informed patients that their new remedy could cure everything from alcoholism to worry (CPC, 1908b).[51] So informed, the patient could avoid the doctor's fee the next time they were ill and procure the nostrum directly from the pharmacist. *Printer's Ink*, the nation's first advertising trade magazine, noted that

> in this way the name of the remedies advertised only to physicians get abroad to the general public … the physician himself … will be the most important medium of advertising at the command of the proprietary manufacturer. In fact, he is that today. (as cited in Sollmann, 1908, p. 25)

In the end, patients who abided by the pamphlet would use phenalgin when acetanilid would do just as well and phenalgin when something else was rationally indicated. As a result, the only thing many

[50] In 1912, a dollar's worth of phenalgin contained four cents' worth of acetanilid (CPC, 1912). Even in cases where the drug was appropriate, the council mourned the pecuniary exploitation caused by a reliance on "ready-made" products. A name-brand deterred doctors from knowing the simple and cheap ingredients that were responsible for the product's efficacy and which could be prescribed in their place by a respectable and competent physician. Given the pseudoscientific verbiage used to obscure the true composition, the council found it "hard to conceive of any one thing that operates more disastrously against scientific therapeutics than the vicious practice of marketing under proprietary names standard and valuable drugs" (Puckner & Hilpert, 1908a, p. 773f).

[51] An alphabetical list of 122 ailments was attached to Antikamnia, but such material was in no way unique.

patients had been relieved of was their money. By pointing out such abuses, the council expected the scales to fall from the medical profession's collective eyes and for the profession to adopt the broader aims of the council as their own.

4.2.2 The fine art of equivocation: Measures and countermeasures

Of the 10 rules instituted by the council, even the brief discussion above identifies several violations and provides an indication of the local environment the council was responding to. Given that rational therapeutics requires the prescription of a specific ingredient for a specific purpose it may seem that only the prohibition of unwarranted therapeutic claims (Rule 6) and the prohibition of unscientific mixtures (Rule 10) are so motivated. However, with proper appreciation of the council's context, every other rule can be seen as promoting rational therapeutics because they were countermeasures to commercial practices that were at odds with best practice.

In order to prevent exploitation, advertising to the public was prohibited whether it was direct (e.g., newspaper ads; Rule 3) or indirect (e.g., listing of indications on the bottle; Rule 4). In the council's eyes, the proliferation of quack remedies was a testament to both the power of advertising and the ignorance of the public. Such strictures were intended to have the doctor intervene on behalf of the patient, where doctors would use their expertise and training to prescribe remedies that had withstood the tests of the laboratory, instead of those that had withstood the tests of the market.

The need to specify ingredients (Rule 1), especially toxic ingredients (Rule 7), on the label was a direct response to the way that manufacturers were circumventing or manipulating doctors' judgment. Sollmann (1908, p. 37) noted, "it is a puzzling psychological problem, but it seems to be a fact, that it is easy to persuade the physician that the same substance acts differently when it is prescribed as 'Antikamnia' and as 'acetanilidum'." Accordingly, any company claiming efficacy over and above what

could be rationally expected from the ingredients had to support their claims with sufficient evidence (Rule 6).

Many drugs adopted the same tactics as phenalgin and claimed that their products were synthetic compounds (not mere mixtures), and thus had medicinal properties over and above what could be achieved by the ingredients. Companies claiming to have created a "new synthetic" were required to include with their submission the proper chemical tests for identity, purity, etc. (Rule 2). The requirement to provide the council with independent tests to establish the identity, purity, and strength of any new compound (Rule 2) is a prerequisite for rational prescription of known ingredients. Moreover, the need for Rule 2 was brought about because manufacturers found it lucrative to claim that their mixtures (which could be cheaply duplicated by any competent pharmacist) were actually sophisticated synthetic chemicals.[52]

The possession of a laboratory allowed the council to independently verify manufacturers' formulae. Given the potential for exposure, reputable firms ceased to make such easily falsifiable claims within a few years of council's establishment (CPC, 1911a). While this was no doubt a positive development in and of itself, it also illustrates a second shortcoming of previous historians' focus on methodological sophistication in isolation: such a view fails to capture the dynamic relation between epistemic aims and the commercial imperatives of medicine. Just as the benefits of castle walls are lessened with the development of siege weapons, the epistemic benefits bestowed by improved methodology only endure as long as they cannot be undermined or circumvented.[53]

---

[52] This was a common advertising strategy. For examples, see CPC (1906b, 1910b), cf. Simmons (1907).

[53] Not exemplified above are: honesty in the origin of ingredients (Rule 5); a prohibition of drug names that did not indicate their main ingredient, especially if they indicated the disease to be treated (e.g., Diabetin, Gonosan, Migrainin, etc.; Rule 8); a requirement to inform the council if patent rights are claimed (Rule 9); and a ban on unscientific and useless mixtures (Rule 10). With the exception of Rule 10, these were rarely cited as reasons to reject applications.

When the council made disclosing accurate information about drugs a requirement, companies responded with counter-countermeasures. Specifically, manufacturers started finding ways of making technically true, but vague statements that would satisfy the letter, but not the spirit, of the rule (CPC, 1914b; Puckner, 1919). Especially when it came to claims related to efficacy (Rule 6), the council noticed

> the skilful indefiniteness that pervades the claims made for [new drugs] which defies scientific refutation. This verbal obscurity is becoming daily more common in the "literature" of firms marketing nostrums … they have reduced equivocation to a fine art. Wherever it was possible to put forward claims by implication rather than by expression this has been done. (Puckner & Hilpert, 1908b, p. 1,706)[54]

Whereas the council sought to eliminate dishonest advertising, manufacturers used words to conceal rather than express, carefully crafting the same message in naively defensible terms. Where the council closed a door to dishonesty, industry searched for an open window.

The matter was complicated by the fact that, much to the council's surprise, doctors did not stop prescribing drugs exposed as fraudulent. Instead, doctors bristled at being told how to practice medicine (Buck, 1909). Companies protested that the council was being too harsh, or had unfairly judged them and falsely impugned their fine product: "so vehement were their protestations and so well simulated were their declarations of Pecksniffian virtue that many physicians were deceived thereby" (CPC, 1908a, p. 467). Whether or not they were deceived, it was at least clear that many physicians paid no heed to the council. An examination of prescriptions three years after the council's inception showed that now roughly half were being written for proprietary concoctions (Motter, 1908, as cited in Flexner, 1910). The next section examines some of the ways that industry had found to sell products.

[54] In the case mentioned, a firm had dropped a drug savaged by the council and then introduced a "new drug" that was virtually identical chemically, but with a distinct name and indistinct claims (cf. Lapius, 1918).

4.3 Humbug is the order of the day: Drug promotion in the Early twentieth century

The tensions caused by commercial imperatives had been recognized long before the council was founded. The vast majority of the efforts by the council were aimed not at *generating* better evidence, but at altering the communication structure of medical knowledge. Before exploring these changes in Section 4.4, it is imperative to consider possible sources of medical knowledge and how these were impacted by the promotional efforts that such reforms were attempting to mitigate.

While the council adopted the mantle of champion of the good and the true, this was not how they were universally seen by the profession they tried to reform. Unsurprisingly, irregular physicians (e.g., homeopaths, osteopaths, etc.) reacted negatively to "being 'bossed' around by 'Simmons' gang'" (CPC, 1918a, p. 1,158; cf. CPC, 1919a). Manufacturers promoted a view that resonated with many practitioners, that the council was a group of "theorists" obsessed with experiments, whereas the proprietary medicine makers were practical men interested in treating the sick (Sollmann, 1908). Additionally, manufacturers attempted to undercut the image of the council in the mind of the public by promulgating the claim that in opposing self-medication they were promoting the interest of the doctors, not patients (CPC, 1915a).

At the turn of the century, doctors could learn about new products from colleagues, medical journals, and advertising. Especially for doctors who practiced outside of the major cities on the east coast, the latter two were particularly prominent. Yet as we will see below, journals and advertising were not as clearly distinct, nor as independent, as they might seem at first blush.

4.3.1 Whose bread I eat, his song I sing: Medical journals and detail men

The ubiquity of the print ad in medical journals provides prima facie evidence that such expenditures provided good returns on investment. Indeed, managers in charge of the economic health of medical

products advised their clients that: “Humbug is the order of the day … It is these efforts that make medicines sell, not so much their intrinsic value. Economy here is no economy at all” (Bickell, 1846, as cited in Gabriel, 2014, p. 57). Even advertising of the more ethically minded houses evolved from subdued, text-based lists of available products, to eye-catching ads laced with illustrations and testimonials from “independent doctors” extolling the product’s merits (McTavish, 2004).

In addition to influencing doctors directly, advertisements in medical journals had indirect effects by altering the content of the journal outside the advertising pages. In 1895 and 1900, the American Medical Association (AMA) judicial council reprimanded *JAMA* editors for putting financial considerations ahead of scientific and ethical obligations. The editor of the *Cleveland Medical Journal* fulminated against “the greed for advertising patronage [which] leads the editor only too often to prostitute his pen and his pages to the advertiser so long as he can secure the coveted revenue” (Forshay, 1900, as cited in Dykstra, 1955, p. 410). It was felt that no periodical could escape the influence of its advertisers. As evidence, the council detailed the practices of journals that ran advertisements disguised as articles (therapeutic notes, commercial news, etc.), noting a tacitly acknowledgement that it was not just the advertising pages that were under the control of sponsors (Salisbury, 1906).

In addition to advertising, journals earned income from reprints. In 1913, for example, *The Army and Navy Medical Record* ran a laudatory article for a cod-liver oil called Waterbury’s Compound (cod-liver oil was a commonly recommended nutrient for diabetics). The company in turn purchased reprints of “One of America’s most valuable preparations” and sent them to doctors across the country. *JAMA* ran a piece reminding doctors that the council's laboratories had shown that if Waterbury’s Compound had any merits, containing cod-liver oil was not amongst them (CPC, 1913a).

In numerous ways, the editorial staff of journals that ran copy for such questionable products defended the interests of their advertisers. When the council found a product to be worthless, other journals defended their patrons in print and cast aspersions on the council (*The Medical Brief*, 1910; *Medical Standard*, 1911). Likewise, editors censored out or refused to print articles that threatened the financial interests of companies that purchased advertising. As one particularly blatant example, consider the curious editorial from the October issue of *Atlanta Journal—Record of Medicine.* The September issue had published a government list of fraudulently marketed products, which included a product advertised in the *Atlanta Journal*. The following edition apologized for including Gray's Glycerine Tonic in the list. The council wondered,

> if 'Gray's Glycerine Tonic' was fraudulently exploited—and the government and the courts have so declared it—why is it necessary for the editor of a medical journal to apologize to his subscribers for having told them so? The only reason that occurs to us is expressed in the caption to this article [Whose bread I eat, his song I sing]. (CPC, 1916a, p. 38)

In addition to print ads, pharmaceutical manufacturers were constantly in search of ways to influence doctors. They sent detail men to visit doctors and inform them of details about the myriad new products being offered by the company. They were not salesmen *per se* because there was never any sale, but they did not come empty handed. Manufactures of remedies such as the digestive tablet Bell-ans were "lavish in [their] distributions of free samples, blotters and other paraphernalia direct to the profession" (CPC, 1915b, 1815).[55] In order to "assist the doctor," drugs often came with pamphlets explaining the theory behind the drug's action and extolling its curative effects (CPC, 1907b, 1914c). Prior to the establishment of the council and its reports, doctors were

> assailed from all sides by the din of the detail man, by the laudatory 'literature' of the advertising pages, the reading-matter, and even the editorial comments of some respectable journals, he was a helpless and easy victim for the skillful proprietor (Sollmann, 1913, p. 6).

[55] It seems that ink blotters were the free pens of the day. It will be interesting to see what happens when doctors no longer literally write prescriptions. For a historic account of detailers see Brody (2008).

4.3.2 One of the most dangerous forms of quackery: The commodification of medical evidence

One of the clearest reasons that changes in evidence cannot be considered separately from the responses of commercial firms to changing epistemic standards is due to the commodification of medical evidence. As doctors begin to rely on a certain form of evidence, the commercial value of that type of evidence increases and commercial firms begin trying to produce it as a means of increasing sales. For example, in the early twentieth century the coin of the evidential realm was testimonials from other doctors.

The testimonials for Buffalo Lithia Water provide a clear example such promotion (CPC, 1914d). A discredited but popular theory held that uric acid was at the root of a number of diseases and given that uric acid was dissolved by lithium, Buffalo Lithia Water was promoted as a new elixir of life. In advertising their products they obtained glowing recommendations from prominent figures such as a former professor of clinical medicine and current vice president of the AMA: "[Buffalo Lithia Water] is strikingly superior to emergency solutions of lithia tablets and pure water, even where the said solution is an exceedingly strong one" (p. 835). It is not clear whether the doctors providing testimonials were guilty of fraud or simply duped, though given the contents of the product and the therapeutic claims made by the manufacturer, the former appears somewhat more likely.

The plausible efficacy of the product can be determined irrespective of whether the uric acid theory of disease was respectable at the time—Buffalo Lithia Water was just bottled water. Lithium was present in such small amounts that chemists estimated a person would have to drink between 150,000 to 225,000 gallons a day to imbibe a therapeutic dose of lithium, and this fact was available to anyone with a modicum of skill in chemistry. Of course, many doctors did not possess any such skills and by soliciting feedback from a fairly large number of doctors, companies were virtually certain to secure a

few favorable testimonials and doctors willing to write up their clinical experiences for publication. The manufacturer was also virtually certain to have such articles published in medical journals from which manufacturers had purchased advertising (CPC, 1914e).[56] One way or the other, Buffalo Lithia Water touted the endorsements of college faculty, the heads of medical societies, and even the Pope's physician.

Buffalo Lithia Water was just one of the worthless products of high repute that were kept afloat commercially. As the nostrum business grew in size it spawned further economic opportunities:

> Whenever a business assumes certain proportions, subsidiary businesses spring up to cater to the needs of the larger enterprise. For some years the nostrum business has grown so large that it has furnished a more or less precarious life for many individuals who have catered to it. There are, for instance, men whose trade it is to obtain testimonials; others, claiming a long string of imposing degrees, will furnish fake reports and bogus analyses; still others issue at irregular intervals publications with high-sounding names which sell editorial indorsement [sic] to the products of concerns such as are willing to pay the price asked. (CPC, 1913a, p. 1,553)

With the rising importance of classic pharmacology in early 1900s, companies began hiring pharmacologists to provide the type of evidence that doctors thought most reliable.[57] These reports were not just used to detail doctors, but found their way into prestigious journals, as when the manufactures of Santogen (a "blood rejuvenator" that was essentially cottage cheese) were able to publish a fraudulent pharmacological report in *Lancet* (CPC, 1914f).

Industrial pharmacologists were barred from professional societies and their work was treated with mixture of contempt and pity as

---

[56] Incurable diseases that had temporary reprieves (e.g., tuberculosis) were especially susceptible to overly optimistic, but commercially valuable, testimonials. As a result, this kind of evidence drew early and heavy criticism from the council: "A year later, the [tuberculosis] patient reposes peacefully under the sod; but the testimonial lives on" (Sollmann, 1908, p. 21). To prevent overenthusiastic and impulsive endorsements, Sollmann advised doctors to put aside for a year any testimonials they planned on submitting.

[57] For a discussion of classical versus clinical pharmacology, see Parascandola (1992, esp. introduction and Chapter 6). For a discussion of the gradual incorporation of pharmacologists into the pharmaceutical industry, see Swann (1988).

> nearly all workers in commercial houses deplore the limitations of their work due to the pressure for financially productive results, and to the necessity of avoiding publications that are inimical to financial interests … [one] need hardly ask for proof that pressure is often put on investigators to supply desirable results. (Hatcher, 1919, as cited in Parascandola, 1992, p. 118)[58]

In general, such information represented an approach to inquiry that views evidence in terms of its economic impact, and as such is willing to manipulate findings in order to increase the economic value of the underlying product. While the standards of what constitutes evidence change, the general strategy does not. Whether it is slanted research produced by a sham laboratory, the clinical report from one of the "stables" of uncritical physicians, or cherry-picked testimonials from elite professionals, "the exploitation of new products by exaggerating their merits and repressing knowledge of failures is one of the most dangerous forms of quackery" (Stewart, 1901, p. 1,177).

4.4 Still in the realms of superstition (1910–1920): Accumulating costly countermeasures

While it would be impossible to draw a sharp line, it is clear that, as the years went on, the council became progressively unsatisfied with the medical profession. The same ploys that had once been exposed with crusading zeal became formulaic and perfunctory. In one case, after detailing a kickback scam for a nonsensical goat-gland goiter cure, the council began a standard condemnation of the *Official Bulletin of the Chicago Medical Society* for advertising the product, but lost the will to editorialize in mid-sentence: "And this sort of pseudo-scientific claptrap is presented to a presumably learned profession through its own official Bulletin—but what's the use of commenting!" (CPC, 1916b, p. 970).

In another instance, the council was confronted with a "new product" that was nothing more than a mixture of aspirin, lithium, and a trademarked name. They lamented: "We had hoped that the

[58] The hackish nature of commercial pharmacology even found its way into popular culture in the Sinclair Lewis' (1925) Pulitzer prize winning novel *Arrowsmith* (see especially Chapter 13 and the description of working for the "pill peddlers" compared to later descriptions of free and unconstrained research at the (fictional) McGurk Institute).

time had passed for reputable houses to employ such time-worn methods, but probably they will not stop so long as physicians encourage them by continuing to use such preparations" (CPC, 1910c, p. 1,803; cf. CPC, 1915c; CPC, 1916c; Edmunds, 1909).[59] While the council continued to agitate for rational therapeutics, when it assigned blame for the persistence of irrationality its rhetoric became progressively less disparaging of manufacturers and increasingly opprobrious of doctors:

> Unfortunately we have seen that the most stirring appeals fall on deaf ears; that those who hear consider that they are not concerned in the abuses; and that the abuses continue unchecked despite the clearest analyses and the most earnest entreaties that can be urged. The time has come to consider whether other measures should not be adopted. (Hatcher, 1916, p. 1,341)

Whereas therapeutic monstrosities had previously appeared as proprietary humbug being foisted on doctors, the continued existence of nostrums became a "slur" or a "sad commentary" on the intelligence of a supposedly learned profession (CPC, 1917a, 1917b).

Commenting on the "educational material" sent out to doctors by manufacturers, the great advocate for residential training, William Osler, wrote in the last contribution of his illustrious career:

> For years the profession has been exploited in this way, until the evil has become unbearable ... We have been altogether too submissive, and have gradually allowed those who should be our willing helpers to dictate terms and to play the role of masters. Far too large a section of the treatment of disease is today controlled by the manufacturing pharmacists, who have enslaved us in a plausible pseudoscience. (Osler, as cited in CPC, 1919, p. 109)

Osler hoped that better education would extricate doctors from their "state of thralldom," a commitment shared by the council. In an effort to reform the field, the council precipitated calls for education reform as part of a long-range effort to create doctors who were familiar with laboratory methods and that would share the council's standard of evidence. In the meantime, the council also

[59] This issue was discussed in some detail in: CPC, 1920. The belief that manufacturers will continue to supply an existing demand was at least in part a result of explicit statements made by manufacturers to the council (CPC, 1918b).

attempted to apply direct pressure by raising the standards to earn their approval and by recruiting others into their sphere of influence so that their decisions were of greater consequence.

Reprising the theme from the previous chapter regarding the effect of the epistemic arms race on the constitution of knowledge, reformers have no interest in creating knowledge so long as what is known has no impact on what is prescribed (i.e., health).[60] In the early years, the council labored under the illusion that poor treatment stemmed merely from the lack of a prominent source of reliable evidence. If they had been primarily interested in knowledge for its own sake, the council's task would have been largely complete in 1907. But as the primary driver of reforms was better patient outcomes, the council began to shift its attention towards the packaging and distribution of knowledge, though, as will be discussed in Section 4.4.2, some of its efforts remained focused on improving the constitution of knowledge.

4.4.1 Nobody who is absolutely worthless gets in: Reforming medical education

Reformers within the AMA, including many on the council, felt that poor education was one of the primary reasons that worthless remedies remained commercially successful after being exposed as fraudulent. This was true in part because of the extremely low entrance standards at many medical schools and because only the elite medical schools trained doctors in laboratory science.[61] Many of the

[60] Again, by "known," I mean to indicate only the maximal epistemic position that an agent could be in given the available evidence at the time, regardless of how many agents actually occupied such a position. For example, it could be said that it was known that Antikamnia worked no better than acetanilid because that is what the best available data justified one in believing. It could be further argued that the extent to which such knowledge was *widely* known is irrelevant to epistemology. In Chapter 7 I will consider a different account of knowledge that is far more applicable to the concerns of medical epistemology.

[61] As an illustration of what could pass for a medical school, consider the "laboratory facilities" at Georgia College of Eclectic Medicine and Surgery. The school was housed in "a building which, in respect to filthy conditions, has few equals, but no superiors, among medical schools. Its anatomy room, containing a single cadaver, is indescribably foul; its chemical 'laboratory' is composed of old tables and a few bottles, without water, drain, lockers, or reagents; the pathological and histological 'laboratory' contains a few dirty slides and three ordinary microscopes" (Flexner, 1910, p. 205). Even more disturbing is the thought that these schools were in the vast

council members reasoned that if doctors were better educated they would be far more likely to appreciate the council's efforts.

In order to appear as a neutral third party, the AMA requested that the Carnegie Foundation for the Advancement of Teaching assess the state of medical education. Abraham Flexner was appointed to survey and evaluate each of the 155 medical schools in the US and Canada. Yet it is more accurate to think of Flexner as a front for reformers at the AMA. The criteria that Flexner used to evaluate the schools were precisely the same as those advocated by the education council of the AMA and the content of the report was substantially similar to an earlier CPC report.[62]

Moreover, the document was more than just a progressive-era equivalent of *The US New and World Report* medical school ranking; it was a systematic treatise on medical education. The first part of the report covered the history of medical training, the economic and professional impacts of over-producing doctors, the necessary qualifications for incoming students, the proper organization of medical schools, the role of state medical boards, and even pressing social questions such as the legitimacy of medical sects, and the education of women and "Negros." The report espoused a revolutionary vision of medical education based on newly emerging laboratory science. It combined soaring rhetoric with caustic assessments, exhorting all concerned to adopt "scientific medicine" and a

---

majority. Flexner described, with a mixture of disdain and eloquence, libraries innocent of books, "laboratories" in proud possession of a single microscope, and dissecting rooms that doubled as chicken coops (pp. 80–89). Still, one might note that even having laboratory space, though foul and ill kept, was a tacit acceptance of the emerging importance of laboratory study for medical education.

[62] This council inspected every medical school in the nation and found that approximately half were acceptable as of 1907 (for more on this report, see Rosen (1983)). This report, while available to Flexner, was unpublished due to AMA ethics codes prohibiting doctors from publicly critiquing other doctors (Starr, 1982). This suggests a likely reason that Flexner, an education specialist, was tasked with the job. For an example of the overlap in evaluative criteria, compare the Flexner report with the criteria advocated by the secretary of the Council on Medical Education for the AMA: Colwell (1909). The extent of Colwell's involvement is unclear, though he may have been Flexner's travel companion for many inspections. While the Flexner Report was ostensibly an independent report, it is clear that the AMA was highly involved. The AMA history identified Colwell as a partner of Flexner's and discussed "their results" (Johnson, 1947).

scientific mindset (even in regular practice).[63] Chapters were devoted both to excoriating the state of education in most institutions and lauding the ideal medical education (e.g., Johns Hopkins, Michigan, and Harvard).[64] It was a standard that few schools could hope to achieve; however, given the superfluity of doctors, Flexner explicitly advised the shuttering of most schools, seeing "no need to make poor doctors—still less to make too many of them"(Flexner, 1910, p. 15).[65]

Though the ideas expressed in Flexner's report were neither wholly original nor clearly decisive in changing medical education, the report still retains its importance as a historical document. It is a stirring and vivid portrait of the consequences of unregulated medical practice, and serves as a reminder that while the council was extolling the virtues of scientific therapeutics, only a handful of schools were teaching doctors pharmacology.[66] In trumpeting the victories of the emerging discipline of pharmacology, Flexner reprised the propaganda for reform.[67] Above all, the document is permeated with two themes that loom large: (1) An encomium to science as the proper foundation for medical

---

[63] For example, "scientific medicine, therefore, has its eyes open; it takes risks continuously; it does not cure defects of knowledge by partisan heat; it is free of dogmatism and open-armed to demonstration from whatever quarter" (Flexner, 1910, p. 53). For discussion of the importance of a scientific approach to clinical practice see (pp. 54–57; 92).

[64] Flexner's alma mater Johns Hopkins served from the beginning as the paradigm of medical education. In preparation to conduct the evaluation, Flexner travelled to Baltimore and spoke with faculty about a scientifically based medical education (Flexner, 1960).

[65] State-by-state recommendations followed the assessments of each of the states' medical schools in Section II. A summary can be found in Section I (Flexner, 1910, pp. 143–155), in which Flexner recommended the elimination of roughly 80% of the existing medical schools.

[66] At the turn of the century, pharmacology was only taught at Johns Hopkins (John Able), Michigan (Arthur Cushney), and Western Reserve (Torald Sollmann). Another 10 schools had added a professor of pharmacology by the end of the decade. See Parascandola (1992, Chapters 3 and 4 for an in-depth discussion of the establishment of academic pharmacology in the US).

[67] Flexner treated pharmacology most thoroughly in his chapter on the ideal course of study during the first and second year (Flexner, 1910, pp. 63–65). After giving a brief history of the accomplishments of pharmacological research, Flexner noted that the modern doctor is threatened both by the tradition of empirics and "the steady bombardment of the unscrupulous manufacturer." He found that what was needed was training in the ability to properly appreciate evidence so that a doctor could both "reject humbug" and accept "really authoritative suggestion." If there was any doubt what authority he had in mind, he directed his readers to *The Propaganda for Reform in Proprietary Medicine* published by the Council of Pharmacy and Chemistry of the American Medical Association.

knowledge; and (2) a repudiation of commercial forces that threatened to sully the doctors' noble reputation as healers of the sick, which, if left unchecked, would ultimately bring medical practice to rack and ruin. Together, it was hoped that such an education would deprive manufacturers of untrained minds susceptible to commercial exploitation.

4.4.2 The committee on therapeutic research: The increasing cost and sophistication of assessing efficacy

Even if successful, the fruits of education reform would take a generation. In the meantime, the council also focused on more immediate issues. When the they were first established, the evidence required to support therapeutic claims was liberal, prohibiting only the grossest abuses and giving the benefit of the doubt to respectable firms. Yet as the council became established, it began to require more exacting evidence than testimonials. As a result, manufacturers of drugs that had previously been included in *New and Nonofficial Remedies* were asked to submit experimental evidence or be omitted from ensuing editions (Sollmann, 1913; for an example, see CPC, 1913b)

A second change was the development of a new way to assess efficacy. As noted above, once the council's laboratory testing made it consequential for companies to misrepresent their ingredients, they began making claims about efficacy where there was no definitive scientific standard beyond what doctors were willing to provide testimonials for. In 1912, the council responded by establishing the Committee on Therapeutic Research under the chairmanship of Torald Sollmann. While many projects were dedicated to identifying drug action, the committee also introduced an innovative method of studying drugs in ways that sought to eliminate personal bias.[68]

[68] For a broader context of methodological development in medical research, see Kaptchuk (1998). While the article is generally enlightening, it fails in several regards concerning the council. First, while Kaptchuk suggested, "the term clinical trial itself was probably coined as late as 1931" (p. 423), the council described its work as

For example, salicylates occurred naturally, but had also been recently synthesized. Companies that had done this claimed that synthetic salicylates were therapeutically more efficacious than natural salicylates, and charged higher prices for their product. These claims were substantiated by the testimony of numerous doctors. The council designed a comparative trial to assess the validity of the claims. The trial began with the committee distributing to clinicians an assortment of numbered—but unlabeled—jars containing either natural or synthetic salicylate. Special care was taken to ensure that neither the patient nor the clinician knew exactly what was being prescribed. Finally, doctors sent back their clinical experiences with the substances so the committee could assess whether assessments differed when doctors were unaware of what they were administering. The design was, in modern

---

performing "clinical trials" (e.g., CPC, 1913c). That said, the meaning of the term has clearly evolved. In the early 1900s, the term applied to a wide assortment of evidence gained from administering drugs to humans instead of animals. The significant distinction for the council was between "haphazard clinical trials" in which doctors tried new drugs on a few patients, and trials like those conducted by the Committee on Therapeutic Research, which were "carefully controlled." The term "controlled" was used more broadly than today, and implied only that the investigator had somehow taken into account confounding factors (Marks, 1997). Second, according to Kaptchuk, Anglo-Americans were not concerned with the effects of suggestion and used placebos to keep patients in the trials to control for the natural history of disease (Kaptchuk, n. 102). While this is true in a broader sense, it does not apply to Sollmann or the council more generally. Sollmann explicitly recommended the use of placebos over a no-treatment control (i.e., natural history) because the "blind test" is "the only method that makes the results purely objective, *really independent of the bias of the observer and the patient*" (Sollmann, 1917, p. 199; emphasis added). Similarly, the council frequently discussed "psychic effects" of inert substances, a concern Kaptchuk claimed entered the literature in the mid-1930s (for a few of many examples, see CPC, 1910d, 1914g, 1915d). Third, Kaptchuk identified the work of Harry Gold as the beginning of the modern era in clinical testing, marking a shift in the American and British understanding of the need for placebos; however, Gold's work merely reflected a continuation of the attitudes of the council. Kaptchuk attributed the first recognition that a sham treatment had an effect to Harry Gold (1931), because he identified an effect of the "encouragement of any new procedure" over and above the effect of spontaneous recovery (p. 426). However, to provide just one example, the council clearly identified this added effect in their campaign to end the indiscriminate use of injections: "The patient is usually interested and impressed by this new, and, to him, mysterious method. There is a psychic element in his reaction to the injection which is not a factor in his reaction to the same drug when given by mouth" (CPC, 1916d, p. 1,451). The council clearly identified that injection has an effect over and above that caused by the physical effects of the drug (if any). Despite the added effect, the council argued that the practice did not justify the danger that accompanied injections, especially in the hands of amateurs. The shared ideas are no coincidence; Gold counted himself as the intellectual heir of Hatcher, council member and student of Torald Sollmann (Gold, 1973). Ultimately, I see no argument for starting the modern era with Gold that does not apply with greater force to the work of the council.

parlance, the first double-blind RCT (Hewlett, 1913).[69] Though such trials were infrequent, they served not only to answer an empirical question, but as methodological exemplars of clinical research that the council expected others to adopt.[70]

4.4.3 Useful drugs: Making rational prescription easy

Similar to their increasing standards for evidence of efficacy, the council also introduced and strengthened its rules over time. When the council began, many of the most lucrative products were what it called "mixtures." Due to advances in chemistry, professional pharmacy had split into manufacturing and dispensing pharmacists (druggists). Traditionally, manufacturing pharmacists supplied the raw ingredients to local pharmacies, where druggists would compound the final product. A doctor's prescription was a formula of botanicals and/or chemicals to be compounded into a pill, tincture, or elixir by the druggist. While there was a small mark-up on raw ingredients (e.g., acetanilid), much larger sums could be garnered by ready-made products (e.g., Antikamnia). A few products required steam machinery or other apparatus not feasibly owned by dispensers; however, manufacturing pharmacists in no way confined themselves to these products alone.

---

[69] Marks (1997) claimed that "as late as 1916 Torald Sollmann was still citing this innovative study as the principle example of the council's work in clinical investigation" (p. 36). However, this study was really one of many Sollmann and the council had organized. The work Marks cited in support of this claim is a summary article of the work of the Committee in the first four years. First, the Hewlett study does not receive special acclaim in the article; moreover, another study discussed in the review article used the same methodology (though its methodology is suppressed in the review—for the full article, see Bastedo (1915). In addition, another double-blind trial was carried out in 1913, using a mixture of oil, cumin, sugar, alcohol, and water as an inert comparator (i.e., a placebo) (CPC, 1913c) and others were carried out subsequently (e.g., CPC, 1917c). While their numbers were not overwhelming, the committee had conducted far more than one clinical trial.

[70] For example, Sollmann (1917) advocated "the blind test" as the only form of clinical evidence that "avoids the pitfalls of clinical observation." To emphasize the novelty of these trials, a canonical history of the emergence of the clinical trial identifies the first collaborative trial as the 1934 serum treatment of pneumonia and the first double-blind trial as the 1931 trial evaluating sanocrysin (Lilienfeld, 1982). Except for formal statistical analysis, the 1913 trial of salicylates had every component of the modern trial. This included random assignment—"the major theoretical innovation traditionally considered to have spawned the modern clinical trial" (Matthews, 1995, p. 128), even though randomization was only formally implemented in 1926, by R.A. Fisher.

Simply put, eliminating mixtures would have been a terrible business decision. By taking a commonly known substance and adding other ingredients, or by reviving a recipe from the therapeutic scrap heap, companies could create "new" products. As Sollmann noted with acetanilid and Antikamnia, doctors would believe that a brand-name product could cure a disease the raw product could not.[71] Alternatively, manufactures could claim their mixtures mitigated the well-known drawbacks of standard remedies, and capitalizing on such credulity was lucrative.[72] While this may have been good business practice, it was contrary to the principles of rational therapeutics.

The council argued that prescribing name brands instead of their ingredients meant that doctors could not control the dosage of the active ingredient; indeed, they may have been completely unaware of what the active ingredient was. As a result, doctors could not apply their experience of one mixture to others that were essentially the same. When there were multiple active ingredients, mixtures necessarily required their administration in a fixed ratio instead of allowing doctors to intentionally prescribe each substance for their patients in the desired amount.[73] Further, doctors would not know what accounted for any observed effect. Above all, the trademarked name of a proprietary mixture sustained a specious mystique. Even in cases where a product was exactly what the doctor intended, the doctor's inability to write their own prescription invariably increased the cost to the patient.

Though the council railed against the unscientific nature of mixtures, it instituted its 1909 ban cautiously at first. The ban represented the first time they had substantially changed their rules. In their initial application of the ban on mixtures, they gave manufacturers the benefit of the doubt, but after

[71] For another example, Migranin manufacturers claimed "In the treatment of migraine with phenacetin or antipyrin, the attack is delayed, while with Migrainin it is usually permanently stayed"—even though Migranin was 90% antipyrin (CPC, 1909a; cf. CPC, 1915e, 1918c).

[72] Examples include, but are by no means limited to, non-nauseating opium (CPC, 1908c, 1911b) and iodine that would not cause iodism in overdose (Puckner & Clark, 1908).

[73] For an example of a mixture with multiple active ingredients, see CPC (1915f). Marks (1997) noted that council discussions showed considerable difficulty in formulating a uniform policy on mixtures with multiple active ingredients.

years of leniency failed to yield a single instance in which a stricter policy would have deprived the profession of a useful drug, the council began requiring definitive evidence of therapeutic value.[74] By taking a harder stance on mixtures, they reduced the number of drugs that intelligent physicians had to consider. While the constructive research on drug action conducted by the council was valuable,

> destructive work is equally as valuable in the drug line. We are still in the realms of superstition … if one goes through these mixtures with a red pencil he will find that it is a minute portion of an iodid, for instance, or a small amount of salicylic acid, that is doing the work and the rest is nonsense. I believe the only official way to show this fact is through the American Medical Association. (Sollmann, 1916, p. 1,442)

The elimination of products was consistent with the council's larger belief that much more therapeutic progress could be made via the simplification of available remedies than multiplication of new ones.

The effort to reduce the number of therapeutic options lay behind a number of the reference books published by the council. Two years after their establishment, the council compiled their reports into *New and Nonofficial Remedies* and made this available "for the trivial price of 25 cents." While *New and Nonofficial Remedies* contained drugs that had been approved by the council, many new drugs were not therapeutic advances. The council noted that the "change of a side-chain makes a new drug; but it may have no more importance than would a change of flavor" (Sollmann, 1913, p. 6).[75] Moreover, while *New and Nonofficial Remedies* supplemented the *Pharmacopia* and *The National Formulary*, the latter two contained hundreds of drugs that were maintained in new editions based on usage, not effectiveness (Sollmann, 1908). With the publication of *Useful Drugs*, the council sought to give a concise list of well-established drugs that could serve as a foundation for doctors' knowledge of therapeutics and an accessible reference for practitioners. Ultimately, the book became a textbook in

[74] For examples of drugs that were rejected or omitted because of the stricter application of rule 10, see CPC (1918 d, 1918e).

[75] The exclusion of unessential modifications was one of the effects of a stricter implementation of useless and unscientific products (Rule 10) in 1918.

many medical schools and, not incidentally, was adopted by many state licensing boards as definitive of what students were required to know to become doctors.[76]

4.4.4 The editor's salary: Changing the finances of medical journals

The enlistment of state boards provides an example of the growing complexity of the council's efforts. Another was their concerted effort to cut off the lifeblood of nostrum promotion: medical journal advertising. While the council had encouraged medical journals to follow its lead and reject any advertising from a product it rejected, high standards had very real economic costs.[77] The prospect of less revenue led a number of smaller journals to resist following *JAMA's* lead. While advertising policy was ultimately an individual editorial decision, retaining disapproved advertising risked the publication's reputation, as the council made no secret of which journals were towing the line and which remained a "gallery of nostrums."[78] Journals that were out of step with the AMA fired back, casting aspersions on the integrity of the council and characterizing the "propaganda for reform" as a self-interested scheme to enhance the AMA's power (The Medical Brief, 1910; Medical Standard, 1911).[79]

In 1912, the council did effectively increase its sway when a number of state journals agreed to abide by the same policy as that of *JAMA*. They gained such support by providing organizational help in securing advertising for state journals in exchange for control over their advertising policy. The following year, the AMA further relieved the burden of individual journals by forming the Bureau of State Journals to vet advertising for state medical societies (Burrow, 1963; Dowling, 1973). This infrastructure was supplemented by liberal amounts of public shaming. Between 1912 and 1915, when

[76] For a list of other examples of council-produced literature, see Puckner (1919).

[77] *JAMA* reported losing $25,000 in advertising revenue in the first year that the policy was in action (CPC, 1907c).

[78] This particular rebuke of the *International Journal of Surgery* appeared in a review of Sulfuryl Monal, but such comments were by no means isolated (cf. Burrow, 1963, esp. pp. 111–4).

[79] While the original name of the council's weekly column was "Pharmacology," this was changed in 1911 to the apposite "Propaganda for Reform."

the council dealt with a particularly reprehensible violation, it also included a list of journals carrying advertising for the products (e.g., CPC, 1912c, 1914e, 1915d). To the extent that the profession followed the council's advice to neither contribute nor subscribe to such journals, at least some of the lure of added revenue from disreputable products was offset by fewer subscriptions. Though such actions brought every state society (except Illinois) under the domain of the council's verdict by 1917, a number of other leading journals continued to advertise fraudulent products as before (Hatcher, 1916; Sollmann, 1913).[80]

This decision to appeal to other organizations (state societies, licensing boards, etc.) reflected a strategic shift for the council. It also entailed a hard-learned lesson: information was not enough. If the goal of the rational therapeutics movement was to improve the care of patients, then the council had to pay as much attention to communication structures as to information being communicated. While it may not count as a "methodological innovation," the publication of *Useful Drugs* had a far greater impact on public health than did publication of the first RCT. Similarly, while some journals adopted stricter advertising standards of their own accord, most did not make the switch until the council had created the infrastructure to eliminate the financial cost of acting ethically. While it is possible that the simultaneous campaign launched by the council awoke the moral conscience of editors, it seems more likely to be an illustration of Upton Sinclair's (1935/1994) insight: "it's difficult to get a man to understand something when his salary depends on not understanding it" (p. 109).

[80] While many top journals continued to run objectionable advertising, others abided by the dictates of the council, especially publications such as *Journal of Pharmacology and Experimental Therapeutics*, where council members served on the editorial board (of the 13 original associate editors, D. Edsall, R.A. Hatcher, R. Hunt, and T. Sollmann were also council members at the time the journal was founded, and one other (C.W. Edmunds) would later become so; for further information on the journal see Parascandola (1992, esp. pp. 136–146)

4.5 The propaganda for reform

The council was created in order to provide the intelligent physician with a source of information about new products independent from "interested manufacturers." Though the widespread use of worthless products was regrettable, it was excusable:

> At first sight it seems disheartening to find that physicians are so easily humbugged. Yet when it is remembered that it is impracticable for physicians either to analyze such products themselves or to go to the expense of having chemists do it for them, it is evident that the fault lies not so much with the physicians as with the conditions that make the exploitation of such frauds possible. (Puckner & Warren, 1910)

However, the council did not seek merely to develop better methods and to replace ignorance with knowledge; it confronted deceit and shaped its rules accordingly. For the first three years, it systematically evaluated an entire class of products and published the sum total of its investigations so that any doctor may have access to these findings. Much to the council's surprise, this effort failed to result in a sea change.

Though the council continued to publish reports, its understanding of the situation changed. They shifted their focus from knowledge production to knowledge distribution, and began to look for ways to amplify their message and to limit the information coming from unreliable sources. In the years between 1910 and 1920, the council engaged in activities aimed at promoting a few established drugs of clear merit. It enlisted influential allies inside and outside the profession and created organizational infrastructure to overcome the financial obstacles that held some journals back from supporting their work.

While the council also developed the method of the "blind test," such research was literally a side-project. They discovered that what is "known" does not matter if it is not acted upon. Ultimately, the dictates of the council mattered only insofar as they informed doctors' beliefs. The collective knowledge of the profession manifested itself in the choices each doctor made, for "with each

prescription he renders a decision whether truth or falsehood shall prevail" (Sollman, 1908, p. 46). Towards this end were directed the main efforts of the council: to focus primarily on its role in developing trial methodology would obscure the day-to-day struggle to lead to lasting change.

Additionally, the details of the council's challenge serve two larger purposes. First, they illustrate that a progressive picture of medical epistemology is seriously incomplete. In place of Matthews' "quest for certainty" or Marks' "progress of experiment," I have argued that medical epistemology is best understood as a competitive interaction in which the commercial drivers of medicine can lead to worse prescription practices. Furthermore, that veritistically oriented developments, such as the council's rules and actions, occur directly in response to, and cannot be understood apart from, the marketing ploys of pharmaceutical manufactures.

Recall that asymmetric races have the following set of features: (1) the reliability of any strategy (once it is employed) typically decreases over time; this is because both (2) opponent responses often attenuate the efficacy of one's strategy and (3) opponents engage in a search process to identify and exploit weaknesses; however, (4) because measures are costly it is often disadvantageous to adopt new strategies until they are necessitated by an opponent; and (5) the process results in the gradual accumulation of costly measures.

The council found that even at the most reputable firms, maximizing shareholder profit conflicted with rational therapeutics. A number of their actions threatened companies' most profitable products, and as a result even reputable manufacturers sought out new promotion techniques that would not run afoul of the council's rules. In response to such countermeasures, the council frequently found that old rules needed to be amended or new rules added (Criterion 1), because manufactures had indeed found ways to circumvent them (Criterion 2). For example, while chemical formulae were easy to verify, companies found that they could make claims regarding efficacy without encountering such

exacting scrutiny. The prime objective of pharmaceutical marketers became promising as much as possible in advertising, without actually saying anything that could be shown to be literally false. This ultimately led the council to begin enforcing stricter requirements for efficacy claims and the development of "the blind test" (see Section 4.4.2). Indeed, as argued in Section 4.2.2, once placed in their historical context, every rule of the council can be seen as a countermeasure to a commercial practice that threatened rational prescription. A similar dynamic of measure–countermeasure was also seen on a smaller scale in the regulation of the chemical constitution of the headache powder Antikamnia in Section 4.2.1.

The actions of the council were not limited to countermeasures. Once they realized that information was not enough, they ceased to be merely reactive. In attempting to reform medical education, the council was not responding to a strategy being employed by manufacturers; they were making an innovative move that they hoped would make advertising inferior products less lucrative (Criterion 3).

One of the less perspicuous aspects of the epistemic arms race is the fact that getting too far ahead is disincentivized (Criterion 4). In some cases this is clear, as when Antikamnia switched to using the more expensive ingredient (phenacetin) to avoid the stigma of listing acetanilid on the drug's label. The case is less apparent with some of the veritistically oriented measures taken by the council. One might object that apart from bad luck, no one could be epistemically worse off by gathering better evidence.

To appreciate the sense in which this is mistaken, consider a bomb maker at the beginning of the Iraq War or a cuckoo parasitizing a naïve species. By the same logic as that followed above, one might say that the insurgent is no worse off using an armor-piercing EFP (see Section 3.2.3) to attack an unarmored Humvee. Similarly, a cuckoo is no less likely to have her small egg rejected if she lays it in

the nest of a species that does not discriminate on the basis of egg size. At worst, the EFP and the smaller egg provide no advantage and, more likely, they yield some small gains. But also recall that tactics are not cost free. The reason that participants in an arms race are disincentivized from getting too far ahead is because the costs outweigh the prospective gains. An EFP requires expensive and sophisticated machine tools to create, while a smaller egg contains fewer nutrients for the maturing bird; these are costs that should not be paid unnecessarily.

In the same way, the council would not have been veritistically better off if it had subjected every product it assessed to "the blind test" for the simple reason that it lacked the fiscal resources and time to run such tests on every product. Given these limitations, conducting blind tests on one drug comes at the expense of some other endeavor. Thus, there were no expected epistemic gains to be had by conducting an RCT on a drug that was a mere mixture, but falsely represented as a new synthetic—time and money were better spent on other endeavors. Yet even with far greater resources, running a clinical trial on these drugs, which were the vast majority of products considered by the council, would not have been the best use of funds given the aims of reformers. With regard to the impact on actual prescription practice, whatever minor gains in knowledge would have been made by running such trials (assuming there were any), a far greater impact on patient outcomes would be had if the funds were put toward more pressing issues, such as the need for reform in medical journals or medical education.

Along the way, the endeavor did become far more extensive and expensive (Criterion 5). The council built a laboratory, published numerous references, began to manage an agency to coordinate the advertising policy of other journals, and lobbied for a number reforms to medical practice. Likewise, the costs of getting a drug on the market began rising as companies began to invest in laboratories of their own and to employ pharmacologists to provide the type of evidence demanded by the council.

In closing, it is worth noting that, contrary to Marks' (1997) account, the council's goal was not "to purge professional therapeutics of commercial influence" (p. 20). They did not aim to drive medicine from the marketplace, but to create a market that rewarded effective products:

> The difficulty has been, and always must be, the fundamental antagonism between objectives that are largely commercial on the one hand and purely scientific on the other. Nevertheless, the Council has always believed and has acted on the belief that there is a possible middle ground wherein the interests of therapeutics would not be injured but would go hand in hand with a commercial development based on enlightened self-interest. (CPC, 1920, p. 1,235)

The goal of reform was not the elimination of private interest from medicine, but the corralling of such interests into epistemically reliable channels. Nevertheless, the council's hope that the more reputable firms would celebrate its work did not transpire; instead, even "the large and old-established firms were not only unwilling to cooperate with the Council, but in many instances exhibited a definite antagonism to the Council's work" (CPC, 1920, p. 1,235). Instead of canalizing market forces in support of rational therapeutics, the council found itself in an epistemic asymmetric arms race.

4.6 Conclusion

This chapter concludes the central argument of the dissertation (viz., that medical epistemology is best understood as an asymmetric arms race). The examination of the council in this chapter emphasizes the fact that recent structures such as CROs are not without precedent. The apocalyptic images from Chapter 1 can now be set in historical contexts. Practices such as ghostwriting are disconcerting, but they are not new. Now, as in the past, many doctors receive their information from pharmaceutical companies and many people end up using worthless or dangerous drugs as a result. In one sense, it is comforting to learn that these are not modern failings. Yet the persistence of the conflict between

patient health and healthy profits cautions against hoping for an enduring solution to the fundamental antagonism.

The final three chapters examine how our thinking must change once medical epistemology is recognized as an asymmetric arms race. One discovery is clear: these battles have been fought before, and there is much to be learned by studying the triumphs and travails of the past century. In the final chapter, I will argue that there are even more lessons to be learned by studying other arms races.

CHAPTER 5

Why Most Sugar Pills are Not Placebos

5.0 Preface

In Chapter 2, I examined the limits of "friction-free epistemology." This chapter explores an example of friction-free epistemology by illustrating that such work can make important contributions, as long as philosophers operate within certain boundaries. Specifically, this chapter argues that the standard philosophical definition of placebos offered by Grünbaum is incompatible with the experimental role they must play in RCTs as articulated by Cartwright.

In order to resolve the discrepancy, I offer a modified account of placebos that respects this role and clarifies why many current medical trials fail to warrant the conclusions they are typically seen as yielding. I then consider recent changes to guidelines for reporting medical trials and show that pessimism over parsing out the cause of "unblinding" is premature. Specifically, using a trial of antidepressants, I show how more sophisticated statistical analyses can explain the source of such effects and serve as an alternative to placebo control.

Though the solution offered here is novel, I also show that the problem is well known and has been so for some time. In Section 5.6, I move beyond the friction-free analysis and explore why such problems have remained unaddressed. In part, the account underscores the fact that the primary objective of most RCTs is to meet regulatory hurdles. However, even when researchers have mounted evidence that inert placebos are insufficient, pharmaceutical companies have been successful in quashing debate.

5.1 Introduction

> *There is a know-nothing school which exhibits a fundamental suspicion of statistics. It is not quite clear why arithmetic and logic somehow become sinister when applied to practical affairs, but it is clear that for some they do. As I see it, statistics is primarily of use for seeing to it that we do not make fools of ourselves in drawing conclusions from our experiments.*
>
> (Lasagna, 1964, p. 97)

It seems rather obvious that when our experiments are potentially confounded, the prudent response is to take better precautions. Yet given the mounting empirical evidence that RCTs designed to be "double-blind" fail to remain so, the CONSORT 2010 guidelines have dropped their recommendation to report whether blinding is successful. The guidelines reason that if a treatment has a significant therapeutic effect, maintenance of the blind may be impossible. I will argue this change represents a step backwards and ultimately results from an inadequate definition of "placebo."

While Cartwright's (2010) analysis illuminated the experimental logic undergirding *ideal* RCTs, worldly vagaries put this logic in tension with the standard philosophical account of placebos offered by Grünbaum (1986). Accordingly, the first tasks of this chapter will be explicating Grünbaum's conception of a placebo and substantiating the claim that, defined as such, placebos are incompatible with Cartwright's ideal. Next, I will provide a *methodological definition of a placebo* that resolves the incompatibility. Finally, I will explore two ways in which RCTs can be altered to bring practice in line with this resolution. The first is to use a control substance that mimics the nontherapeutic effects of the drug (e.g., side effects). The second is to use an inert substance, but to attempt to control for the nontherapeutic effects statistically. Regardless of how this incompatibility is resolved, clarifying the nature of the tension is salutary. It both identifies the situations where medical trials fail to warrant the conclusions they are standardly taken to demonstrate and suggests how they might be amended so as to yield more reliable results.

5.2 Grünbaum's Placebo

Behind Grünbaum's (1986) work on placebos was an internecine debate in the mental health community over the efficacy of various schools of psychotherapy. With the example of psychotherapy in the foreground, it was clear that a placebo is not just any treatment that relies on psychological effects,

and it is similarly flawed to identify placebos with inert substances. For instance, water, salt, and sugar are paragons of therapeutically inert substances; however, while inert generically, they can be effective treatments for dehydration, hyponatremia, and hypoglycemia, respectively. Such modest examples elucidate Grünbaum's central insight: that a treatment is a placebo with respect to some disease (D). More specifically, Grünbaum divides a treatment as follows: A therapy (t) is explained by a theory (ψ) as being composed of: (1) the characteristic factors (F) that are purportedly responsible for improvement and (2) the incidental factors (C) that are not (see Figure 1). For example, suppose a therapeutic theory holds that the chemical fluoxetine is the proper treatment for depression. Since Prozac and Sarafem are chemically identical (they are both fluoxetine), they have identical characteristic factors. However, while they also share some incidental factors (e.g. they are both manufactured by Eli Lilly), there are others they do not share (e.g. Sarafem pills are pink and purple while Prozac pills are green and white; Sarafem is roughly 15 times more expensive than Prozac).

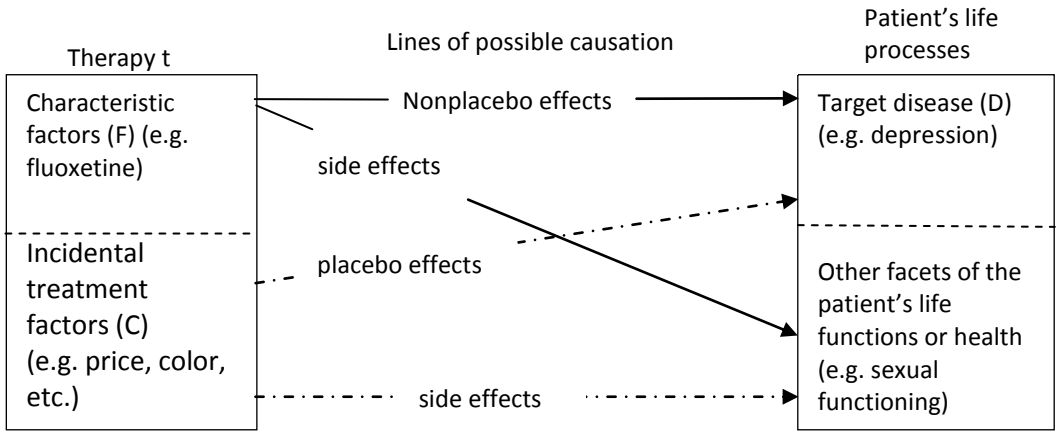

Figure 1: Grünbaum's Model of the Treatment Process
An example of Grünbaum's model used to specify drug/placebo effects.

Next, Grünbaum divides the effects of the treatment on the patient (whom he calls the "victim") into (i) the intended effects on D and (ii) other aspects of the patient's life (i.e. side effects). Staying with the example above, Prozac might cause an improvement in mood as intended, but unintentionally cause sexual dysfunction.

A therapy (t) is "an intentional placebo" (a placebo intended for treatment) for a victim (V) suffering from D treated by a practitioner (P) if: (1) none of the F positively affect D; (2) P believes the first condition is true; (3) P believes that C will positively affect D for V; and (4) P generally allows V to believe that t has remedial efficacy for D because of F. More plainly, condition 1 entails that a sugar pill is a placebo as long as the patient was not helped by the sugar. Conditions 2–4 specify that the patient, but not the doctor, believes the treatment is effective because of some characteristic factor. In these terms, most sugar pills are placebos.

5.3. The Crystal Palace

To highlight potential problems with Grünbaum's (1986) conception of a placebo, consider the following thought experiment. Suppose crystal healing is the standard form of treatment, but we were inclined to doubt the efficacy of crystals. Crystal healers (and thus, ex hypothesi, the population at large) believe that sandstone is the appropriate treatment for some disorder. To assuage our concern, a crystal healer performs an RCT. In the experimental group, sandstone is used to treat the patients. In the control group, amethyst (which has no expected effect) is used instead. The amethyst and the sandstone are enclosed in a device that allows for the appropriate skin contact while preventing the patient and the healer from seeing which is being used. In follow-up evaluations, while both groups have improved considerably, the treatment group has significantly better outcomes. Suppose further that this finding is

repeatedly confirmed. Healers conclude that this constitutes scientific evidence to support the widely accepted claims of efficacy.

Such trials would meet the general standards of EBM and even the requirements of the FDA. Standard practice allows that the difference in outcomes between patients who received sandstone and patients who received amethyst can be attributed to the effectiveness of the sandstone treatment. Amethyst also meets Grünbaum's (1986) definition of a placebo control (a placebo intended for a trial) iff (1) none of its characteristic factors positively affect D; (2) P believes that amethyst is harmless for V with D; and (3) P wishes to know whether any of the observed improvement can be attributed to the characteristic factors.

Yet we can grant each of these conditions, while maintaining that the test was not properly controlled. Suppose I hypothesized that it is possible to determine patients' group assignment during the trial. To support this, I produce evidence that patients, doctors, and independent evaluators are all able to make this discrimination at levels above chance. There are two possible explanations therein. It could be because all of these groups correctly believe that sandstone works. For example, subjects who recovered reason that, since they recovered, they must have been treated with sandstone. Subjects who did not recover reason in the same way (*mutatis mutandis*) and conclude they have been treated with amethyst. Alternatively, it could be the case that after the groups were randomized, amethyst did not mimic "nontherapeutic" aspects (e.g., side effects) of sandstone effectively, thus allowing subjects to discriminate between treatments based on properties incidental to treatment.

In line with the latter hypothesis, suppose that when sandstone is rubbed on the body it causes abrasions, whereas amethyst does not. Further, suppose I show that patients guessed their treatment assignment on the basis of abrasions, not improvement. Finally, suppose we find the following: When you statistically control for which group patients *believed* they were in, patients were no better off if

they were *actually exposed* to sandstone. In light of these facts, I believe a reasonable conclusion is that sandstone was not an effective treatment despite its apparent superiority. Its only independent effect was to cause abrasions. The presence of abrasions signaled to subjects that they were receiving the culturally accepted medical practice, which in turn resulted in improvement via an expectancy effect. Because the abrasions were a side effect (i.e., they were an effect of the drug on the patient's life processes, in contrast to an aspect of the drug itself), such an expectancy effect does not realize Grünbaum's conception for either a placebo effect or a nonplacebo effect (see Figure 1). Let us call this "the sandstone effect."[81]

5.4 The placebo control revisited

The methodological definition of a placebo is determined by the underlying logic of ideal RCTs. As articulated by Cartwright and Hardie (2012), the ideal RCT is a manifestation of Mill's method of difference. Patients are randomized into two groups. The control group receives precisely the same care as the treatment group except for x, where x is what is being evaluated for efficacy. Given that patients receiving treatment t improve and the only difference between t and t* is x, x must be the cause of the improvement.

Faced with the situation above, it seems that we must either give up the standard definition of placebo, or conclude that an RCT is insufficient for establishing efficacy. In the remainder of the paper I will parse out the consequences for the former. In place of the traditional definition, I will propose *the methodological definition* of a placebo. In essence, this definition turns "placebo" into a success term. As a first pass, a treatment t* is a placebo in an RCT testing t for D iff it plays the methodological role

[81] To the extent that doctors alter their practice or their assessment of patients based on which group the doctor believes the patient is in, there will be a corresponding sandstone effect for doctors (since they may see through *their* side of the double-blind). For simplicity, I will focus on the effects on patients, but this account applies equally well to doctors.

required to determine whether t is a treatment for D. As an example, let us consider arthroscopic knee surgery for osteoarthritis of the knee.

Moseley et al. (2002) randomized patients into one of three groups. In the incision group, a patient was locally anesthetized and several incisions were made, but nothing more. The second group also had their joint lavaged (washed out) and the final group had their knee lavaged and débrided (surgeons removed damaged tissue). If the actual procedure was not performed, the surgeon still went through the motions of asking for the instrument, manipulating the patient's leg, and even splashing saline to simulate the sounds of lavage. In this procedure a number of factors are clearly incidental: the fact that Moseley was a celebrated surgeon, the sound of splashing water as the joint is rinsed, etc. Other factors are less clearly categorized.

Surely, the group we *consider* to be the placebo group will depend on what our theory ψ determines to be the characteristic factors of t, but much can hang on such designations. Surgeons diverge in whether they attribute the therapeutic effect to the débridement, the lavage, or both (all agree that the incision is incidental). According to ψ_1 the débridement is the characteristic factor and the lavage is merely an incidental prelude. In contrast, ψ_2 supposes that the lavage is the characteristic factor of the procedure and the débridement is both incidental and otiose. Finally, ψ_3 posits that both the lavage and the débridement are characteristic factors that each make a contribution to the therapeutic effect. The correct theory of arthroscopic knee surgery classifies as *characteristic factors* all and only the causally relevant factors of the treatment that exert a therapeutic effect independent of the patient's expectations. The choice is not conventional; the other theories are false.

The definition of a characteristic factor as noted above implies that any factor that does not independently improve therapeutic outcomes is an incidental one. In the sandstone parable, the amethyst group improved (though to a lesser extent than the sandstone group); however, according to

the society's prevailing theory, mere exposure to amethyst outside of the healing ritual would not have resulted in any therapeutic effect. Accordingly, any therapeutic effect that did occur must be the result of the patient's expectations.

Beyond their therapeutic outcomes, characteristic factors may also cause side effects (i.e., effects unrelated to the disorder). Side effects might be caused independent of patient belief, as when the incision during knee surgery results in soreness or, contra Grünbaum, as a result of expectancy effects (Kirsch, 2010). For the sake of clarity I will use "expectancy effects$_t$" for therapeutic effects and "expectancy effects$_{se}$" for side effects (see Figure 2).[82]

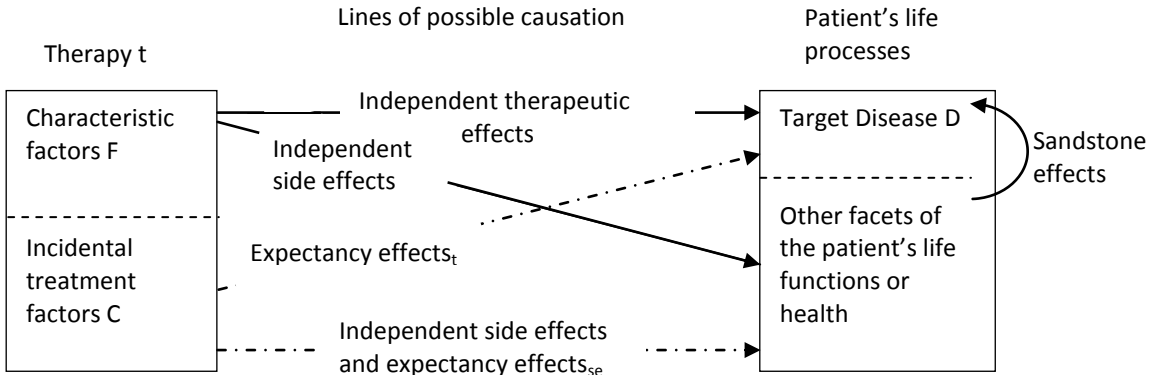

Figure 2: Grünbaum's Model of the Treatment Process Amended

Grünbaum's basic model stays intact; however, expectancy effects are not limited to therapeutic effects and side effects may impact D.

Returning to arthroscopic surgery and the subsequent controversy, we are now in a position to spell out the methodological role required to determine whether t is a treatment for D. With regards to ψ_1, only the lavage group serves as a suitable control. If the débridement group were compared with the incision group, a superior efficacy would not have established that the débridement alone was

[82] It i's worth noting that because side effects can be beneficial (e.g., better cardiovascular health is a side effect of treating depression with aerobic exercise), these are not necessarily "nocebo effects" (negative placebo effects). Moreover, since there can be negative expectancy effects on D, nocebo effects are not limited to side effects.

responsible for the improvement. According to ψ_1, the lavage group contains all of the incidental factors C of t and thus controls for any expectancy effects caused by C.

As it turns out, Moseley et al. (2002) found no difference between the lavage and the débridement groups. Nevertheless, this does not show the procedure is ineffective; it is possible that the lavage had an independent effect on osteoarthritis. If so, the lavage was not a placebo control, but an effective therapy. Such a possibility can be assessed by examining the difference between the lavage and the incision group. This comparison is complicated by the fact that these groups had different levels of pain shortly after the surgery, allowing for the possibility of a sandstone effect. Fortunately, Moseley et al. (2002) assessed patients' beliefs about which group they were in, as well as their expectations for a successful recovery, and found no differences in either. Thus, the incision treatment served as a placebo control for the lavage treatment. Furthermore, as Moseley found no difference between the incision and the lavage, the latter was also a placebo for débridement.[83]

In summary, the methodological definition of a placebo is as follows: t* is a placebo in an RCT testing t for D iff (1) t* methodologically controls for the expectancy effects$_t$ caused by t;[84] (2) none of the incidental factors of t* have an independent therapeutic effect; and either (3a) t* produces all of the same side effects as t; or (3b) despite the failure of (3a), patients' beliefs about which treatment they are receiving is the same for both t and t*. It is worth noting a number of nonstandard implications of this definition. First, whether t* succeeds in serving as a placebo for t should be supported by evidence. Second, contra Grünbaum, it is not enough for the placebo to simply lack the characteristic factors of t; it must share the incidental factors. Moreover, a placebo can, and in many cases must, cause side effects independent of any expectancy effects. For example, if patients in the control group did not

[83] Technically, it could be the case that the incision had some independent effect on osteoarthritis. Though nothing rules this out, I am not aware of anyone who takes this position.

[84] Essentially, t* contains all and only the incidental factors C of t. This condition is similar to Howick's (2012, p. 82) requirement for a "legitimate placebo."

have to heal from the incision, the procedure would fail to serve as a placebo control *in virtue of the fact* that it caused no side effects. I think that one can, if one wishes, maintain that a "placebo effect" that is independent of a patient's expectation violates one's concept of a placebo (viz., placebos should be inert); however, as alluded to above, I do not think that one can maintain that position and continue to have placebos serve the methodological role they currently occupy in medical research.

5.5 An alternative to placebos

The view outlined above cuts against current trends in medical research. Recently, in light of the fact that blinding consistently fails in practice, the CONSORT 2010 guidelines retracted their previous recommendation for trials to test to see if blinding is successful, claiming, "We now believe the interpretation of a test of its [blinding] success to be dubious, if not impossible" (Schulz et al., 2010, p. 1,145). The guidelines claim that because blinding might fail due to patient improvement, such a test may simply reflect the efficacy of the drug instead of the methodological failure of the trial. Recall that the logic of RCTs is that to attribute the efficacy to the drug the only difference must be characteristic factors. Thus, the CONSORT 2010 guidelines essentially confront a possible failure of the central assumption required by the statistical analysis used to assess efficacy with the unsubstantiated hope that everything will work out.

Fundamentally, this problem is an outgrowth of the fact that inert substances often fail to serve as adequate placebos. If RCTs used a placebo in the sense I propose, no such ambiguity would arise, as the therapeutic effect would be the only possible cause of unblinding. However, since most drugs have side effects, so too would most placebos. Nevertheless, for either practical or ethical reasons, it may not be possible to give patients a substance that causes all of the known negative effects to be caused by treatment, but without any of the characteristic factors hypothesized to be beneficial. An alternative to

the methodological control is to administer an inert substance (viz., not a placebo) and attempt to control for these effects statistically.

Consider a trial on antidepressants carried out by the NIMH that used an inert substance in the control group.[85] Given the methodological definition of a placebo, whether the trial was in fact placebo-controlled must be substantiated. Since the control group was given a lactose pill that was identical in appearance to the antidepressant, the design of the trial provides *prima facie* evidence that the placebo group controlled for all of the incidental factors of antidepressant treatment without containing any independent therapeutic effects. However, while the first two conditions of a placebo were met, patients in the antidepressant group experienced far more side effects. Moreover, patients were able to guess which group they were in at levels far above chance. In short, the trial was not a placebo-controlled experiment and the possibility that the observed superiority is due to a sandstone effect cannot be ruled out.

The ability of participants to correctly identify which group they were in does not attest that the drug was ineffective, but the standard practice of comparing group means is not justified. Instead, a more complicated procedure should be employed to achieve statistically what a placebo is intended to achieve methodologically. The null hypothesis is that there is no therapeutic effect of the drug. Accordingly, if patients were able to identify which group they were in on the basis of therapeutic improvement, then the null hypothesis is false (the treatment is effective) and the logic of the RCT is not threatened (Howick, 2012). In line with such reasoning, proper analysis must ascertain the cause of patients' beliefs about group assignments. If, as in the antidepressant trial, patients' guesses were the result of differences in side effects, a sandstone effect cannot be ruled out and we cannot reject the null hypothesis.

[85] In what follows, I will describe my own results in less than full detail. However, a more in-depth analysis can be found in the technical appendix and additional information provided upon request.

In such cases there is a solution that can be achieved using the type of data already collected in many trials. Instead of testing group differences, structural equation modeling or regression analysis can be used to test a proposed model of symptomatic improvement. Specifically, statistical analysis can be used to ascertain whether one variable acts as a mediator between two other variables (Holmbeck, 1997; MacKinnon et al., 2002).

First, suppose that we have three variables, A (treatment), B (belief about group assignment), and C (symptomatic improvement), that are each highly correlated, and we wish to determine whether A has an independent effect on C. This results in a model structure with three paths, as in Figure 3.

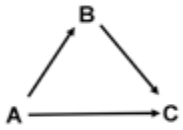

Figure 3: Proposed Causal Model

In the model above arrows indicate lines of hypothesized lines of causation. Accordingly, A is thought to causally effect B and C; likewise, B is thought to causally effect C

We have evidence of the sandstone effect if four conditions hold: (1) A is a significant predictor of C; (2) A is a significant predictor of B; (3) B is a significant predictor of C after controlling for effect of A on C; and (4) A is a significantly poorer predictor of C when B is controlled for.

To get a feel for the model, consider the crystal palace thought experiment. To assess crystal healing, let A be the actual treatment, B be which treatment the patient believes they are receiving, and C be the degree of symptomatic improvement. In the proposed situation, sandstone causes an expectancy effect, so there will be a relation between A and C. Further, since patients used abrasions to discriminate between "real crystal healing" and placebo crystals, there will be a relationship between A and B. In addition, all of the variance in improvement is explained by people's confidence that they are receiving "real crystal healing," so there will be an effect of B on C (since crystals do not actually heal, controlling for their contribution to healing (A) will not change the effect of B on C). Finally, the effect of

crystals on healing disappears when we control for belief, so A is a significantly worse predictor of C when B is controlled for. Thus, all four conditions are met.

With this more sophisticated analysis we can see that the pessimism of CONSORT 2010 is unfounded. With sufficient data, it is entirely possible to identify the cause of unblinding and subsequently tease apart the direct effect of the treatment from the expectancy outcome caused by side effects. In the trial of antidepressants, it turns out that patients' guesses are driven by side effects and not improvement. Patients who received antidepressants improved more than the control group did, and patients who believed they were receiving antidepressants improved more compared to patients who believed they were receiving placebos. Crucially, once what a patient believes is accounted for, there is no added benefit to actually receiving antidepressants. In short, this trial, once properly controlled for statistically, provided no evidence that antidepressants were an effective treatment. The apparent superiority of antidepressants in this case is caused by the sandstone effect. Given that the patients (and doctors) could determine which group the patients were in by the presence of side effects, the lactose pill does not meet the methodological criteria to be considered a placebo; however, we might reasonably call trials analyzed in this fashion *statistically controlled RCTs*.

5.6 The role of placebos in the arms race

To this point I have concerned myself solely with friction-free epistemology; that is, what sort of conclusions is one warranted in drawing given the evidence routinely contained in an RCT. I have thus far argued that the standard manner of analyzing RCTs does not, even in ideal cases, warrant the conclusions that are typically drawn from them. However, there is another question one might ask, namely what steps have thus far been taken to address this issue. If science was exclusively driven by veritistic aims, one might think that as soon as the problem was identified, researchers would attempt

to find a solution, and that once found, the superior methodology would become standard practice. This is not what has happened.

5.6.1 In the beginning

The first indication that the failure of the double-blind detailed above was (or should have been) dealt with is how quickly such an issue arose in the medical literature. In this section I will review some of the earliest indications of the failure of the double-blind. Unlike later disputes, these debates occur entirely within the community of researchers carrying out RCTs.[86] The fact that such discussions date back to the first trials of antidepressants indicates that the failure of the blind was an obvious phenomenon that, at least initially, presented a problem to overcome, rather than a threat.[87]

As noted in Section 5.1 the first published report on imipramine was by Kuhn (1958), but this consisted of unblinded clinical observations where such an issues could not arise. Yet with regard to the very first double-blind trial of imipramine conducted with out-patients, Ball and Kiloh (1959) remarked: "With regard to preknowledge of what the patient was receiving before the code was broken, we had the usual hints from the side-effects that occurred" (p. 1,053). No further action was taken, but the comment itself indicates that it was already common knowledge that the double-blind was more of an ideal than a reality. Letemendia and Harris (1959) reported that since such deficiencies had been noted informally, they would attempt to assess these claims empirically. Though they found that it did not alter the nurses' assessment of psychotic patients hospitalized for at least 15 years, their recognition of the implication of their results is so clear that they can be quoted here in full:

---

[86] This is not to suggest that there were not significant disputes over RCT methodology between clinicians and trialists; indeed there were (see Marks, 1997). The point here is that the particular problem of how to maintain the blind was a problem addressed within those who conducted trials.

[87] Moncrieff, Wessely, and Hardy (1998) suggest that ethical considerations spoke against using active placebos, but this assertion is unsubstantiated.

> The argument for imposing double-blind conditions is that if the observers are not told which patients are receiving the drug they cannot be influenced by prejudice for or against it. They cannot however be prevented from making guesses ... They may guess either more correctly or more wrongly than chance. This is the possibility realised in the present trial, and one might suspect the usual state of affairs. In this case, the double- blind procedure would appear to be invalidated to a greater or lesser extent depending on the degree of departure from chance. (pp. 45–46)

Though they do not provide any sound ways of dealing with the problem, it is enough to observe that problems with maintaining the blind were not only suspected, but had already been shown experimentally.

During the late 1950s and throughout the 1960s, determining what should be standard procedure for assessing efficacy was still being worked out and expectancy effects were a subject of experimental investigation (Nash, 1962). Concern for such effects can be seen, for example, in Heaton-Ward (1962, as cited in Leyburn, 1967). In his evaluation of the effect of nialamide on children with Down syndrome, Heaton-Ward made it appear to the psychiatrists conducting the trial that a double-blind crossover placebo trial would occur. Each bottle used in the trial was labeled with one of two numbers. Halfway through the trial, the numbers on the bottles were switched (but not the actual treatments, which remained constant throughout) to make it appear that the crossover phase was now beginning. In the first half of the trial, the children receiving nialamide were rated as significantly improved. Similarly, in the second half the children ostensibly receiving nialamide (but who were actually maintained on placebo) where seen as having improved. Heaton-Ward concluded that despite the double-blind design, the side effects had unblinded the raters in the first half of the trial and this bias carried over when the raters believed the treatments had been switched. While this trial was not conducted with antidepressants, it was relayed as a cautionary tale in an early review of antidepressant trials published in the *Lancet* (Leyburn, 1967), and so was highly visible. Not surprisingly, Leyburn called for more trials using active placebos.

What is perhaps even more impressive than early identification of the problem is the rapidity with which a potential solution was found. Almost immediately, Daneman (1961) conducted a trial in which he attempted to control for anticholinergic side effects (e.g., dry mouth) by using atropine as an active placebo. Further, use of such a control was not an isolated occurrence. An active placebo was used in at least five other trials during the 1960s (Friedman et al., 1966; Hollister et al., 1964; Uhlenhuth & Park, 1963; Weintraub & Aronson, 1963; Wilson et al., 1963). However, reformers did not see this solution in time to affect policy (e.g., Lasagna, 1960). In 1963, just as these studies were being published, the FDA codified the double-blind procedure using an inactive placebo as the experimental design required to establish efficacy. With the inert placebo endorsed as the industry standard, only a few additional studies employed active placebos (Friedman, 1975; Hussain, 1970); nevertheless, it is safe to say that problems with achieving a truly double-blind trial are not a recent discovery.

5.6.2 The problem resurfaces

In 1982, R. Thomson published a paper in *The British Journal of Psychiatry* that revisited studies from Morris & Beck's (1974) influential meta-analysis, widely regarded as definitive in establishing the effectiveness of antidepressants. Thomson noted that whereas roughly two-thirds of the trials that used an inert placebo judged antidepressants favorably, only one in seven of those that employed atropine (as an active placebo) found the drug to be superior. Judging from the lack of citations of Thomson's paper after its publication, this article was virtually ignored. A few other articles commenting on the issue of blinding did arise (Brownell & Strunkard, 1982; Oxtoby, Jones, & Robinson, 1989), but these too met with either dismissiveness or silence.[88]

[88] For example, a reviewer for Oxtoby, Jones and Robinson's (1989) initial submission commented: "The luddites rise again!", as if noting that most studies do not include a manipulation check is anti-scientific.

A sudden interest in trial methodology arose in the late 1980s amongst psychologists and was most likely the direct result of the meteoric success of Prozac. A spate of books and articles were published on the topic between 1989 and 1997 (Antonuccio et al., 1995; Bystritsky & Waikar, 1994; Fisher & Greenberg, 1989; Fisher & Greenberg, 1993; Fisher & Greenberg, 1997; Greenberg & Fisher, 1989; Greenberg et al., 1992; Greenberg et al., 1994; Margraf et al., 1991; White et al., 1992; Young, 1996). These articles revived interest in Thomson (1982) and generally questioned the notion that antidepressants worked, given that they were based on the integrity of the blind in RCTs. Though these studies tended to cite one another, they failed to arouse a more general concern.

5.6.3 The discrediting of active placebos

Finally, in the late 1990s a few articles were published (Kirsh & Saperstien, 1998; Moncrief, Wesley, & Hardy, 1998) that got picked up in widely read popular press (e.g., Blakeslee, 1998; Day, 1998; Enserink, 1999; Horgan, 1999; Talbot 2000). Not incidentally, a more general defense of antidepressants and inert placebo trials was finally published by authors favorable to industry (Quitkin, 1999; Quitkin et al., 2000). After reanalyzing the same data, they suggested that the breaks in the blind were the inevitable consequence of an efficacious product and concluded that "the active placebo theory gains no support from these data" (Quitkin et al., 2000, p. 332).

Ostensibly, this reanalysis was conducted because recent media attention might discourage people from seeking treatment. Indeed the article was framed as something that would help doctors who were unfamiliar with the research literature in "facilitating an informed discussion of the pros and cons of antidepressant treatment and questions about true drug efficacy" (Quitkin et al., 2000, p. 328). However, rather than a genuine attempt to correct a popular misconception, the analysis appears to be

a rearguard action in response to bad press.[89] Though the Quitkin et al. (2000) analysis is methodologically dubious, the article served the rhetorical purposes of discrediting the opposition.[90] Research psychiatrists, almost without exception, found the Quitkin et al. (2000) article definitive and effectively ended the debate. Most crucially, the Quitkin et al. results were cited in support of eliminating the requirement to assess blinding in the CONSORT 2010 guidelines for how to conduct RCTs.

5.6 Conclusion

This chapter first outlined the limits of friction-free epistemology by concerning itself with the logic of inference irrespective of commercial forces that effect the actual practice of medical research. I argued

[89] A close inspection of Quitkin et al.'s (2000) article reveals a number of interpretive choices that consistently favor industry. For example, for one of the trials in the meta-analysis, Quitkin et al. claimed the results favor the use of antidepressants even though the original study (Hollister et al., 1964) found no difference. Examining the original article shows that Quitkin et al. ignore non-significant results for the primary measure. The secondary measure they refer to contains seven subscales and compares the placebo to two different drugs. For five of the subscales the differences are non-significant. In one case, the placebo is inferior to both drugs, and in the last case the placebo is superior to one of the drugs. Hollister et al. conclude: "Although each drug was superior to the control treatment in some respect, most differences were small and inconsistent between drugs" (p. 374). In contrast, Quitkin et al. glossed these results as a trial that "supports antidepressant efficacy." In addition to questionable interpretative choices, Quitkin et al. obtain their meta-analytic results by transforming the original data in each trial into a binary "responders/nonresponders" variable. No indication is given of a principled rule for the transformation and my attempts to reconstruct a uniform principle have not been able to produce one. In some cases, Quitkin et al. considered the top two possible ratings as a responder (Friedman, 1975; Friedman et al., 1966; Hussain, 1970), while in another, they counted only the top rating as responders (Daneman, 1961). Similarly, no principle can be created by using the qualitative descriptions consistently across studies as there are inconsistencies in how qualitative descriptions are transformed. For example, Quitkin et al. coded "moderately improved" as a responder in one study (Friedman, 1975) and a non-responder in a second (Friedman et al., 1966). Even if a principle could be given for the decisions discussed above, in each case the data are grouped in such a way as to create the maximum difference between responders and nonresponders. Any other principle will decrease, if not eliminate, the apparent effect of antidepressants. Indeed, the creation of the binary variable itself is suspect as it reduces the granularity of the data. By creating a binary scale, Quitkin et al. lost information. It seems that it is due to this blurring of the data that they obtained significant results from data that, untransformed, shows no difference between the drug and placebo (e.g., compare Quitkin et al. to Friedman, 1975).

[90] For a few examples of articles citing Quitkin et al.'s (2000) study as definitive refutation of the concerns about active placebos, see Young (2001), Hollon, Thase, and Markowitz (2002), and Rupniak (2003).

that even in the ideal case, an inert substance cannot serve the methodological role it is tasked with, and thus the standard meaning of "placebo" stands in need of revision. All too often trials are called double-blind or placebo-controlled on the basis of their design. There is now significant evidence that for any condition susceptible to expectancy effects treated by a drug that causes side effects, providing an inert substance to the control group will fail to warrant the assumptions that are standardly used to analyze the data. Despite such evidence, the pharmaceutical industry has been effective at forestalling and quashing the use of active placebos, and has now succeeded in removing even recommendations to assess whether the blind was maintained in a trial.

Correcting this design flaw is one potential next move for medical reformers. The same rationale that motivated the introduction of inert substances into trial design can be marshaled to argue for a truly adequate placebo (i.e., one that mimics side effects). Nevertheless, ethical or practical concerns may militate against the use of such placebos. In such cases, statistical controls can be used to attenuate methodological shortcomings. In either event, trials must be described not by their intended design, but by what conditions actually obtain, and reporting guidelines should be changed accordingly. If researchers accepted such methodological strictures, then it is overwhelmingly likely they will conclude that most sugar pills are not placebos.

CHAPTER 6

The Problem of Intransigently Biased Agents

This chapter was co-written with Justin Bruner

6.0 Preface

In Chapter 2 I argued that applied questions in medical epistemology could not be answered without taking into account commercial imperatives. In this chapter, I will make an initial attempt to incorporate non-truth-seeking agents into an epistemically pure community. Extending a formal model of network epistemology pioneered by Zollman, I show that an intransigently biased agent prevents the community from ever converging to the truth. I next explore two solutions to this problem, including a novel procedure for endogenous network formation in which agents choose whom to trust.

While this series of models captures aspects of problems faced by the medical community, the problems explored here are static: neither epistemically motivated agents nor intransigently biased agents change their strategy for influencing the community. While these models are a step forward in including commercial imperatives, the static nature of the agents' strategies leaves something to be desired. This further complication will be addressed in the final chapter.

6.1 Introduction

The emergence of social epistemology has provided both a new range of philosophical problems and new formal tools to address them. A salient aspect of this new approach has been the examination of how information is shared among agents in a group. Surprisingly, features that would intuitively seem to be epistemic virtues, such as free exchange of information, can turn out to inhibit the group in question from acquiring true beliefs (Zollman, 2007). More generally, instead of one optimal communication structure, epistemic virtue depends crucially on the particular problem confronting the group (Zollman, 2013). This paper will consider a formal model of one issue that increasingly confronts diverse areas of scientific inquiry: *the problem of intransigently biased agents*.

Previous studies have assumed that research is conducted by agents who, broadly speaking, are interested in discovering the truth (e.g., Alexander, 2013). However, there are broad swaths of science in which those who are financially backing research do so with the express aim of promoting a claim regardless of underlying facts. For example, tobacco companies funded work that delayed the establishment of a causal link between second-hand smoke and lung cancer. The potential consequences of regulation and taxes for the energy sector led the fossil fuel industry to fund studies that controverted the reality of anthropogenic climate change (Oreskes & Conway, 2010). Chemical companies fund research that minimizes the effects of exposure to toxic substances in order to reduce their legal liability (Elliott, 2011). And so on. The presence of financial interests in these domains fundamentally alters the incentives that drive scientific inquiry. Consequently, epistemically motivated inquirers in these areas must contend with intransigently biased agents.

To illustrate this problem, I will first examine the use of diethylstilbestrol (DES) as a prophylactic for miscarriage. Next, I will review networks of agents confronting the bandit problem as a model of social learning. The static nature of communication in these networks exacerbates the problem of intransigently biased agents, suggesting that if agents are allowed to choose whom to trust, they might be able to avoid manipulation. This chapter uncovers that such freedom can render biased agents ineffective at misleading the community.

6.2 DES: Four decades of intransigence

In the decades before antibiotics revolutionized medical care, endocrinology was in ascendance. However, like many advances, endocrinology brought with it excess enthusiasm in both a host of legitimate products and in similar-sounding quack remedies. It is somewhere in the penumbra of legitimacy that we find DES. The synthetic estrogen began with an excellent pedigree. In 1939, the

British Medical Research Council reported favorably for the use of DES in several conditions related to menstruation and menopause. Because no patent was sought, any drug company that wished could manufacture and market DES.[91]

In 1941, 12 companies gained FDA approval to use DES to ameliorate the symptoms of menopause, and seven years later it was approved as a prophylactic for miscarriage. Yet not all of the research on DES was favorable. Pervasive side effects, as well as animal studies demonstrating that DES was carcinogenic, led to some at the AMA to recommend that it not be recognized for general use, characterizing contemporary practices as "overzealous … indiscriminate and excessive" (Stoddard, 1945, as cited in Dutton, 1988, p. 47).

Amongst early proponents for the prevention of miscarriage were Harvard professors Olive and George Smith, whose research formed the substantive basis for FDA approval. Though well respected, the work was not without its critics, and within five years four separate (methodologically superior) studies had shown that DES was ineffective. Unfortunately, the FDA had come to the conclusion that it lacked the legal authority to remove ineffective products from the market. Although the FDA was explicitly awarded such legal authority in 1962, it would take until 1971 before officials concluded that DES was contraindicated for pregnant mothers.

Meanwhile, roughly 100,000 prescriptions/year were written throughout the 1960s. By the end of its use at least 3% of the nations' children had been exposed to DES *in utero,* in addition to the millions of mothers that had ingested it (Meyers, 1983). Ultimately, it fell from favor in large part due to the actions of women and men who brought public attention to the increase in cancer, deformed genitalia, and fertility problems caused by their *in utero* exposure. However, given that many of these problems were known or suspected from the start, a perennial question has been: What explains the

[91] This account is heavily indebted to the excellent scholarship of Dutton (1988).

continued use of an ineffective and dangerous drug? Amongst the possible answers, which will be considered below, are: studies published in medical journals, expert opinion, doctors' and their colleagues' experience, and information provided by pharmaceutical companies.

We have already seen that DES was not supported by the medical literature. By 1954, over 2,000 women had already participated in four RCTs, all of which failed to support efficacy, and the largest of which showed that DES *increased* miscarriages.[92] As for experts, while the Smiths never recanted, their position became increasingly isolated. Internal memos document that the DES manufactures themselves were aware that use of DES was being rejected by the medical elite (Dutton, 1988).

Thirdly, there is the experience of the doctors themselves. Given that DES exacerbated the problems it was prescribed to ameliorate, doctors' experience should lead them to the conclusion that DES was ineffective at best. However, this is overly hasty for two reasons. First, doctors might simply encounter a random string of live births and mistake these for drug efficacy; second, doctors might be so sold on an intervention that failures are perceived as successes (e.g. DES caused a premature birth, but the doctor assumes the baby would not have been miscarried without DES). Yet the former reason would not explain such widespread uses and the latter possibility only pushes the question back to where these doctors' enthusiasm came from. While some fervor might be attributed to the progress of medicine in general, the lion's share can be found in the information provided by pharmaceutical companies.

One common and influential source of information for doctors was the *Physicians' Desk Reference* (*PDR*). The information contained in the *PDR* was submitted by the manufacturer and then sent out to doctors free of charge. In 1960, over half of busy doctors consulted it daily (Dutton, 1988).

[92] See Bamigboye and Morris (2003) for a retrospective analysis of the available literature.

Beginning in 1947, DES was listed as indicated for "habitual or threatened abortions" with no mention of any disconfirming evidence until 1969, when the indication was dropped and a strong warning against use in pregnancy was added.

Pharmaceutical marketing was both passive and active, and all of it sang the praises of DES. Magazine ads ranged from relatively subdued pieces listing only the claims approved by the FDA to garish ads recommending DES for all pregnancies (see Langston, 2010). More active marketing involved company spokesmen (detailers) whose job was to visit doctors and keep them up-to-date on the companies' products. Corporate memos clarified the approach that detailers were to take to doctors: "Tell 'em again and again and again—Tell 'em till they're sold and stay sold" (as cited in Dutton, 1988, p. 58).

Many doctors would deny that such sources affected them, claiming to be men and women of science moved by reason, not the same tricks used to sell soaps (Avorn, Chen, & Hartley, 1982). Yet in the case of DES, no other source besides advertising appears to be a viable candidate for explaining such widespread and enduring use. In the face of detailers "telling doctors" and "telling them again," many doctors were indeed sold and stayed sold. Moreover, it seems that nothing that the doctors told the detailer could change their mind. Even a biased agent would have been moved to reconsider their position if they were in search of the truth, but detailers and other sources of information such as the *PDR* were not just biased: they were intransigently biased.

6.3 Network structure and the bandit problem

I now move to a precise formal framework that can help to better explain the influence that biased agents have on a group of epistemically pure agents. In particular, I examine a network of individuals that are all confronted by the so-called bandit problem, a situation in which one is presented with two

slot machines and must determine which to play. Zollman (2007) suggested that this is analogous to a doctor determining which of two medications to administer.

Doctors are modeled as Bayesian learners, who update their beliefs when presented with new evidence, and are myopic in the sense that they simply administer the drug they believe is more efficacious. Moreover, there is no guarantee that an individual doctor will correctly identify the more efficacious drug. Consider the following scenario: a doctor has observed 5,566 successes upon administering drug A 10,000 times, and only 10 successes upon administering drug B 20 times. In this case, our agent will believe drug A is superior, but clearly, since comparatively little is known about drug B, the optimal long-run strategy may include prescribing B to gain more information. The myopic doctors considered in the course of this section, however, will only begin to prescribe B if the success rate of A falls under 50%.[93]

In our model, the doctors do not know the true success rates of drugs A and B. In each interval doctors administer to their N patients the drug they believe to be superior—where each patient has probability p_A (or p_B) of recovering—and record what percentage recover. Doctors are embedded in a social network and treat results obtained by their neighbors on par with their own experience. As Figure 4 indicates, epistemic agents are represented as nodes in a graph, and those nodes connected by a line are said to be "neighbors."

With the society of knowers in view, we can now ask some interesting questions; chief amongst them, how should the group communicate in order to maximize the likelihood that every member will learn which drug is superior? While agents in the maximally connected graph reach consensus more quickly, those in the cycle are more likely to reach a true consensus (Zollman, 2007). This counterintuitive finding occurs because, as connection density increases, the entire group becomes

[93] Myopia is plausible in several cases, including where a doctor feels ethically prohibited from giving a patient a drug perceived to be worse just to increase the doctor's confidence that the drug is inferior.

more likely to be converted from the superior option by a chance wave of poor results. By contrast, the cycle promotes situations in which the group as a whole stays undecided for longer, and there is at least one member collecting data on each option—a phenomenon Zollman (2010) called "transient epistemic diversity."

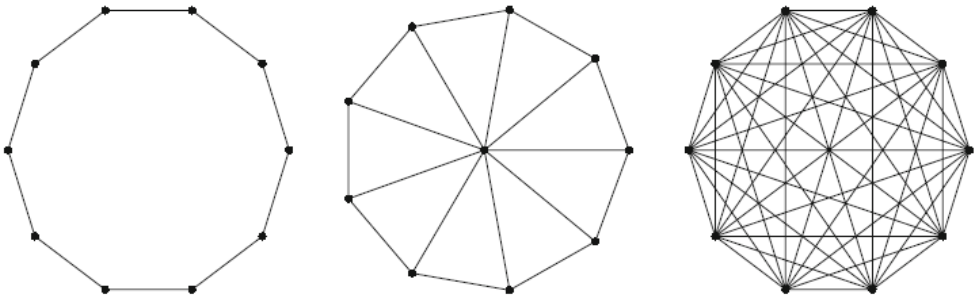

Figure 4: Idealized Epistemic Communities

Each node is an agent and each line represents a two-way communication channel between the agents. We refer to these three canonical structures as "the cycle" (left), "the wheel" (middle), and "the complete graph" (right).

I demonstrate in the present chapter that these results only hold so long as agents are epistemically pure. Generally speaking, an impure agent is one that is interested in convincing the group of a view irrespective of the truth. In the medical field, this would be an agent, such as a pharmaceutical company, that attempts to encourage doctors to use a drug irrespective of which drug is more efficacious. In my simulations, epistemically impure agents administer only their favored drug and the results they obtain are produced by a biased distribution. Thus, if the actual probability of success is 56% and the bias is 10%, the impure agent reports data as if the probability of success is 66%. Specifically, their data comes from a binomial distribution with a mean of 56%+*b*, where *b* is the

strength of the bias. This is my attempt to capture, in my idealized model of medical epistemology, the fact that pharmaceutical manufacturers find numerous ways to subtly bias their results.[94]

I focus primarily on the "worst-case scenarios" in which the pharmaceutical company promotes the inferior drug and is connected to all doctors.[95] In terms of Zollman's canonical network structures, it is a wheel with the biased agent at the hub. Briefly, the exact set-up is as follows. Agents are randomly assigned beliefs regarding the two available drugs. Doctors, as well as the biased agent, administer the drug they believe to be most efficacious. This generates N data points, and all share their data with those they are connected to. Doctors then update their beliefs in a fashion outlined by Zollman (2010), and subsequently repeat this process.

6.4 The impossibility of sustained convergence to the truth

Consider the case in which drug A is successful with a probability of 0.51 and the pharmaceutical company's drug (B) is slightly inferior ($p_B = 0.5$). Assume all doctors begin with true beliefs regarding both drugs. Given this belief profile, all will immediately begin administering A to their patients. We have convergence in the short run, but not in the long run. This is due to the bias of the pharmaceutical company (assume the bias is 0.03). Since the pharmaceutical company is the only one conducting research on B, it alone influences the doctors' perceptions about it. As the doctors' impressions of B improve, one of the doctors will "crossover" and begin to administer B. By doing so, this doctor is now running their own *unbiased* experiment (viz., they are now flipping their own coin, which provides unbiased information to the doctor and their neighbors). This in turn helps to mitigate the influence the pharmaceutical company has on everyone the doctor is connected to, including the doctor him- or

[94] For a non-exhaustive list, see Safer (2002).
[95] Given manufacturers' ability to organize "educational events" and fund "key opinion leaders," maximum connectivity is a reasonable approximation of reality (Elliott, 2010; cf. Krimsky, 2003).

herself. Thus, if the doctor and two of their neighbors both switch over to the pharmaceutical company's drug, the combined results of their experiments are sufficient to mitigate the influence of the pharmaceutical company and move back to the superior drug. Yet when none of the doctors investigate B, the only information they receive about the drug, once again, comes from the biased pharmaceutical company. We now see why convergence to the superior drug for a sustained amount of time is impossible in Zollman-type models with intransigently biased agents.

In order to quantify this effect, I look to the last 1,000 rounds of a 2,000-round simulation and determine how frequently the best drug was used. The result is that six doctors arranged on the wheel use the superior drug 42% of the time (see Figure 5). Interestingly, this number increases as more connections are added to the network. In the complete network, doctors utilize the superior drug 63% of the time. Thus, in contrast to Zollman (2007), the more connections there are, the more likely the network as a whole is to adopt the more efficacious treatment. The reasons for this should be obvious. When doctors are better connected to each other, fewer doctors have to spend their time debunking the biased results because the unbiased results are more widely broadcast.

Though more connected networks provide a defense against intransigently biased agents, nothing short of eternal vigilance is required of the community, which must constantly devote members to investigate the less successful drug. This keeps the biased agent at bay, but is surely a second-best solution. I argue that what is primarily driving the need for eternal vigilance is the fact that experimental results from one agent are taken just as seriously as experimental results from another. If individual doctors could learn that the pharmaceutical company is severely biased, doctors may begin to discount the company's results. The above models are *static*, meaning that epistemic agents must listen to everyone they are connected to. If this assumption is relaxed and we consider *dynamic networks* in which connections can change, our doctors may learn to ignore the pharmaceutical company. I will now consider endogenous

network formation and show that if individuals have some control over whom they listen to, then for a wide variety of parameters, the pharmaceutical company is unlikely to draw doctors away from the most efficacious drug.

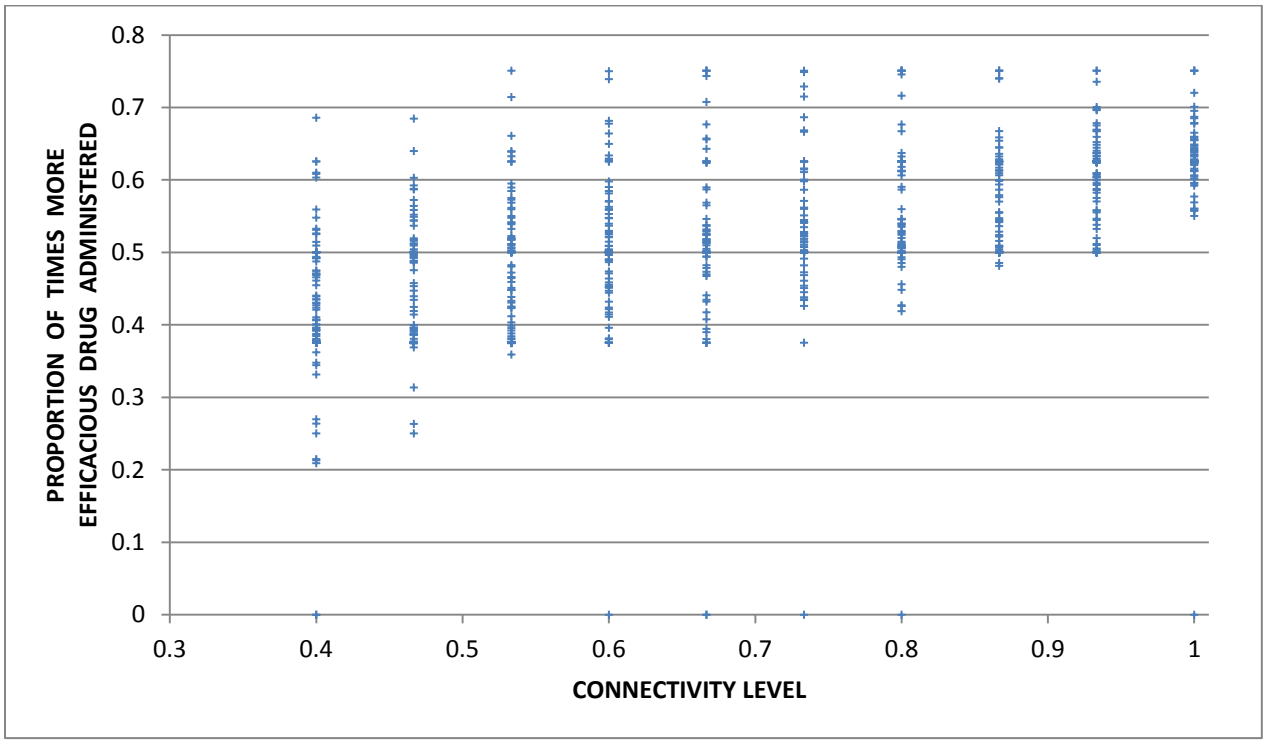

Figure 5: Simulations on a Static Network with a Central Intransigently Biased Agent

Connectivity level refers to the proportion of possible connections between doctors (e.g. the wheel has 6/15 possible connections). Each "+" represents one simulation. As the connectivity level increases (i.e., connections are added), the doctors prescribe the superior drug more often.

6.5 Choosing your neighbors: Endogenous network formation

Modeling network formation is an active area of research in a number of disparate fields.[96] Unfortunately, none of the canonical models can be appropriately applied to the epistemic community in question because the agents are continuously generating data. As in my earlier model, a doctor *i* is

[96] See Jackson (2005) for an overview.

connected to doctor *j* if *i* is somehow influenced by the experimental findings of *j*. However, in this model, network connections now vary continuously and are no longer symmetric, meaning that *i* can be strongly connected to *j*, while *j* is only weakly connected to *i*. In this case, *i* is strongly influenced by *j* while *j* is only slightly influenced by *i*. Similar arrangements no doubt do occur, such as when the work of a senior scientist is very influential on a junior scientist, but this influence is not reciprocal. In general, *j* strengthens their connection to *i* if *i*'s experimental findings are in line with *j*'s current subjective beliefs.[97] Likewise, *j* weakens their connection to *i* the more that *i*'s experimental findings seem to clash with *j*'s beliefs. Making this precise is difficult, and highlights why many models of endogenous network formation used in economics and sociology are not applicable when thinking about such an epistemic network.

I instead present a novel model of endogenous network formation that replicates basic hypothesis testing inside an epistemic community of agents that are continuously experimenting. Consider a network of *D* doctors. Each doctor has *D+1* bins (one for each doctor and one for the pharmaceutical company) that initially have anywhere from zero to 100 balls in them. Let B_i be the vector $< b_{i1}, b_{i2}, \ldots, b_{iD+1} >$ where bi_1 is the number of balls in agent *i*'s first bin. How strongly connected agent *i* is to agent *j* is determined by the proportion $b_{ij} \Big/ \sum_{k \in D+1} b_{ik}$.[98] This connectedness determines how much weight *i* puts on the experimental findings of *j* (call this w_{ij}). Agent *i* updates their beliefs regarding drug A as in Zollman (2010), except that the results are weighted as follows:

$$P(\text{drug A works} \mid \text{agent } j \text{ has s successes in N trials}) = (\alpha + w_{ij}\, s) / (\alpha + \beta + w_{ij}\, N)$$

[97] For example, if *i* thinks that a drug is efficacious 50% of the time and *j* reports that the drug worked in 51/100 trials, then *i* would strengthen their connection to *j* in future rounds.

[98] For example, suppose there were only two other doctors. The first agent would have three bins (one is for the agent him- or herself). Suppose agent one started with 10 balls in their own bin, 20 balls in agent two's bin, and 30 balls in agent three's bin. When the results from the next round come in agent one will weight their own results by multiplying them by 1/6, agent two's results will be multiplied by 2/6, and agent three's results will be multiplied by 3/6. Weights are normalized so that they always add up to 1.

where α and β are the agent's values from the previous round. *Ceteris paribus*, the more balls agent *i* has in their *j*th bin, the more connected they are to agent *j* and thus the larger impact agent *j* has on *i*'s beliefs. Individuals adjust their connections in the following fashion. Upon receipt of N data points from agent *j*, agent *i* conducts a one-sample t-test based on their subjective beliefs. Let t_{ij} be the t-score agent *i* assigns to agent *j*'s experimental results in round *r*. The number of balls in b_{ij} is then updated by the following equation:

$$b_{ij}(r+1) = b_{ij}(r) + f\left(t_{ij}\right), where\ f(x) = \Lambda(1.96 - |x|) \text{ and } \Lambda > 0.$$

Where b_{ij}(r) is the number of balls in agent *i*'s *j*th bin at round *r*. Thus, a t-score with an absolute value of less than 1.96 results in an increase in the number of balls in the bin, while a t-score with an absolute value exceeding 1.96 results in a decline. How the strength of connection to *j* is affected can of course only be determined if we take into account the change in all bins. One intuitive property this update rule satisfies is the following: if you are connected to two individuals and they repeatedly provide you with the same evidence, then in the long run you should expect to be equally connected to these two individuals. One's initial connectivity "washes out" in the end.

The inclusion of network formation has drastic effects. One common outcome is for all doctors in the community to heavily discount the pharmaceutical company's experimental data. In this case, none of the doctors administer the inferior drug and all have minimal connections to the drug company. The biased agent is effectively quashed, thereby allowing doctors to converge on the superior drug. Less desirable arrangements are also possible. In some scenarios, a minority of agents listen to both their fellow doctors and the pharmaceutical company. The level of connection these agents have to the company does not completely dissipate because the company had a hand in shaping their perception of the drug.

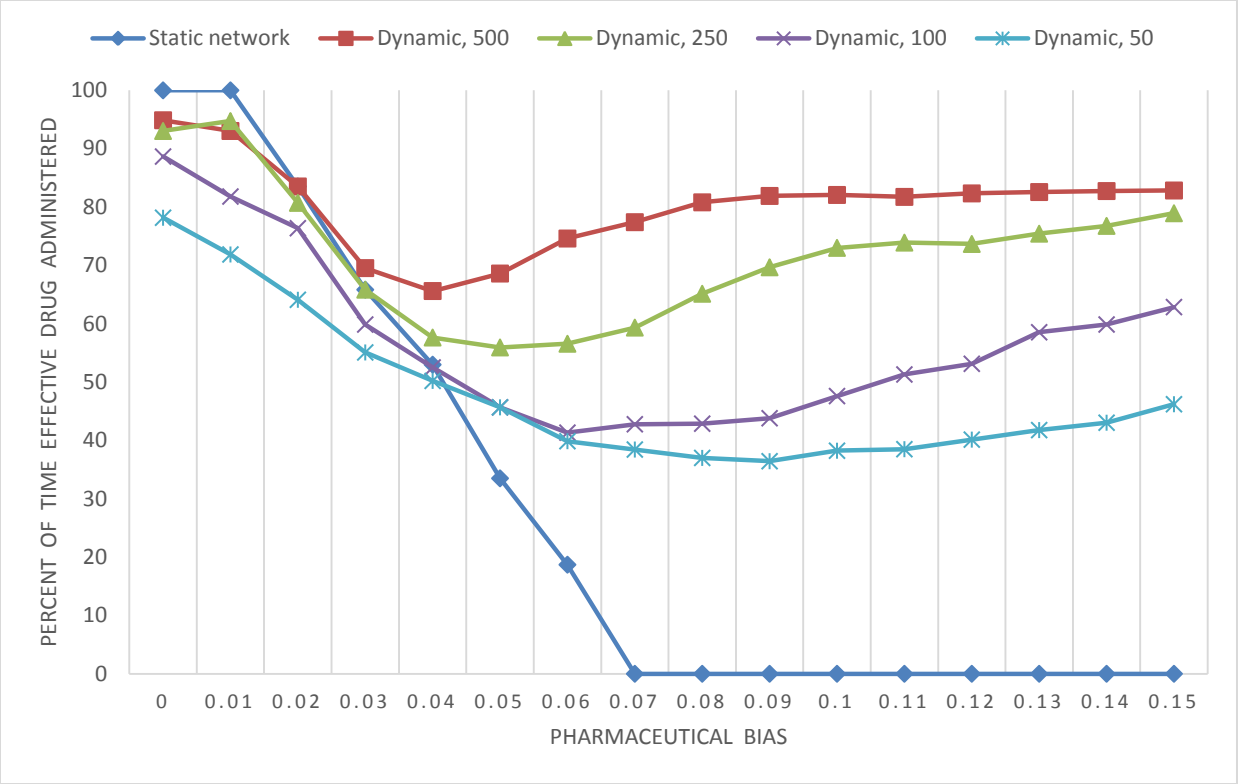

Figure 6: Simulations of Dynamic Networks

The graph displays proportion of times the more efficacious drug is administered in the last 1,000 rounds (y-axis) for different levels of *b* (pharmaceutical bias), ranging from 0 to .15 (x-axis) as well as various levels of N (i.e., 500, 250, 100, 50)

The company's biased experimental results are thus not seen as particularly unusual, since they are in some sense already reflected in these doctors' subjective beliefs.

By and large, however, a dynamic network helps the community to better identify the superior drug. For example, in the static network with $p_A = 0.51$, $p_B = 0.50$ and b = 0.08, doctors almost never come to prescribe the superior drug. In contrast, the superior drug is prescribed 80% of the time in the dynamic network. In general, dynamic networks are much more resistant to the influence of intransigently biased agents than static networks, and Figure 6 drives this point home quite nicely.

Two variables are primarily responsible for ensuring that the more effective drug is taken up by the population: N and *b*. As N increases, the community becomes more likely to converge on the better drug. Surprisingly, convergence on the superior drug is also more probable when the company is highly biased. All else being equal, if the bias is outlandish, then even a small number of trials will be able to alert the community that something is awry. Introducing biased data can influence honest agents, but lies have to be subtle enough to go undetected. In a dynamic network, agents can simply stop listening if the bias becomes apparent.

6.6 The problem of intransigently biased agents and epistemic clarity

The problem posed by intransigently biased agents can be alleviated if these agents learn to identify and trust good informants. We have seen that this is not possible in a static network, since by decree individuals cannot come to ignore their neighbors, thereby allowing a biased agent to mislead the community. Furthermore, Zollman's finding that "in small finite groups, the best graphs are minimally connected" (2013, p. 25) fails to obtain with the introduction of biased agents. Instead of promoting a virtuous transient epistemic diversity, the lack of communication forces sparsely connected agents to duplicate the debunking work—if they are able to resist the biased agent at all.

The introduction of the network formation rule yields desirable results. While other update schemes may be superior, this simple rule prevents agents from being manipulated by a highly biased pharmaceutical company. It creates a point at which increasing the bias in one's results merely makes it easier to be identified as untrustworthy. Even in cases where the pharmaceutical company retains some influence with most doctors, groups virtually never converge to the wrong drug, and under most circumstances reviewed here, they prescribe the right drug more often than not. Indeed, one common

result is that every doctor gives no weight to the pharmaceutical company and roughly equal weight to everyone else.

Returning to the DES case, doctors most closely approximate agents in the static wheel. Each doctor was in contact with a limited number of colleagues, but maintained contact with the pharmaceutical company via advertisements, the *PDR*, and interactions with detailers. Thus, despite the experimental evidence, elite opinion, and the doctor's own experiences, use of DES continued apace. The models considered here suggest two possible responses: increase the number of connections or learn to ignore biased agents (e.g., stop meeting with detailers).

It might be suggested that the DES disaster could have been averted if doctors had simply been trained to pay attention to the results of RCTs as is currently recommended by the EBM movement. As described above, these results showed a lack of efficacy, but note that this just pushes the problem back. As doctors have become more influenced by research, pharmaceutical companies have come to spend an increasing amount of their marketing budget on biased trials (Angell, 2004). A number of meta-analyses have found a large correlation between positive results and industry funding (Bekelman, Li, & Gross, 2003). Rochon et al. (1994) found that 56/56 comparison trials funded by manufacturers of nonsteroidal anti-inflammatory drugs for arthritis concluded that the funder's product was as good as or better than the comparison drug. While this was particularly egregious, it has been estimated that between 89% and 98% of trials yield results favorable to the company that funded the research (Cho & Bero, 1996).

Given the severity of this problem, some commentators have suggested that pharmaceutical companies be prohibited from conducting this type of research. An alternative to such a fundamental change in the structure of scientific practice is to better exercise epistemic discrimination. Though it is rare, official bodies have occasionally considered devaluing the epistemic weight accorded to industry-

funded studies, a proposal the British advisory agency NICE considered, but ultimately rejected. The present analysis suggests that something like our network-formation rule may be preferable to the current practice of treating all equally well-designed trials as equivalent regardless of their source.

6.7 Conclusion

This chapter has provided an example of how the insights of Chapter 2 might be incorporated into a social epistemological framework, yet the discussion has already highlighted a shortcoming of these models. The narrative has been framed in terms of strategies that truth-seeking agents could adopt, and what intransigently biased agents might do in response; however, the models explored here do not permit agents to adopt different strategies. In the final chapter, I situate the arms race into a fully social epistemology and consider what implications this has for medical epistemology.

CHAPTER 7

Embrace the Suck

Lessons from other Arms Races

7.0 Preface

> *[Embrace the suck] is a quip, but it's also a command and an encyclopedic insight: it's gone bad. This is a tough situation. You wouldn't be here if it weren't a broken, challenging, dangerous situation. And that's the suck. To put it in context, a historical context, General von Clausewitz talks about friction, how everything is going to go wrong in war … But that's the suck. And the idea behind embrace the suck is that, look, the situation is bad. Now let's deal with it.*
>
> Austin Bay (Iraq war veteran)

In Chapter 2 I argued that there is a significant difference between error and manipulation, and that previous philosophical accounts (friction-free epistemology) fail to address manipulation. By examining both the current state of medical research in Chapter 1 and the historical cases in Chapters 2 and 3, I have shown that manipulation is an entrenched and pervasive impediment to generating sound medical knowledge. The previous chapter provided a specific example of a social epistemological approach to understanding medical research. It is a case that illustrates how philosophical morals can change when commercial forces are incorporated.

In this chapter I will argue that, over and above this particular example, the conceptual framework offered by social epistemology provides the best understanding of medical epistemology. Having done so, I will conclude the dissertation by considering what larger vistas become available by seeing medical epistemology as an asymmetric arms race. I will argue that methodological rigor cannot be equated with epistemic reliability, that sources of information must be included in evaluating evidence, and that commercially unfavorable results should be accorded a disproportionately high weight. I will also argue that when considering a change in epistemic practices, the goal should be a policy could not be easily circumvented (i.e., robustness). I conclude by examining the consequences of abandoning universal standards of evidential reliability. I acknowledge that such a result is undesirable; however, once one acknowledges that medical epistemology is fundamentally antagonistic in character, the result is as inescapable as it is unsatisfying. Nevertheless, considering medical epistemology

provides purchase on how to understand the problems, and guidance as to what approaches are likely to result in improvements. In the words of American soldiers, embrace the suck. The situation is bad, now let's deal with it.

7.1 Is this philosophy?

As described in Chapter 1, there are several threats to the attainment of reliable medical knowledge, including disease-mongering, a low bar for obtaining patents/FDA approval, an imbalanced confluence of interests, the manipulation of doctors' medical judgment, pharmaceutical companies assuming the form but not the character of scientific rigor, and the emergence of contract research organizations. However, to reprise an objection from Chapter 2, why should a philosopher worry about advertising?

A response that I feel particularly inclined to give is to agree with Karl Popper that philosophy is the study of problems. In order to understand the work of canonical philosophers, Popper argues that it is necessary to study the history of science and mathematics because traditional philosophical problems arise out of "urgent and concrete problems, problems which they found could not be dismissed" (Popper, 1963, p. 73). According to Popper's account, Plato's philosophy stems from wrestling with the irrationality of the square root of two from within a Pythagorean framework (that asserted essence of reality is numerical). Kant's *Critique of Pure Reason* is an attempt to understand how it was possible for Newton to have discovered the truths of physics. These are presumably genuine philosophical works, if any there be, and what they show is that "*genuine philosophical problems are always rooted in urgent problems outside of philosophy* … What matters is not methods or techniques, but a sensitivity to problems and a consuming passion for them; or as the Greeks said, the gift of wonder" (Popper, 1963, p. 72, italics in original).

This is not to pretend that the forgoing is on par with the great works of philosophy, but merely to suggest that, like these prior works, it is motivated by an urgent problem that falls outside the discipline. Briefly, it is the problem of intransigently biased agents. More generally, we consider science to be our best way of knowing about the world, but does science function the same in contexts of heavy industry funding? If not, how does it function and should it still be considered our most reliable source of knowledge? I have argued that the answers to these questions are: no; that it functions as an asymmetric arms race; and below I will argue that the answer to the last question is no as well (at least on some occasions). However, some may think that Popper's definition is too broad and so the fact that this project fits within it merely begs the question.

If so, perhaps one could characterize it as a project in second philosophy. According to Maddy, the second philosopher "begins from common sense perception and proceeds from there to systematic observation, active experimentation, theory formation and testing, *working all the while to assess, correct, and improve her method as she goes*" (Maddy, 2007, p. 2, emphasis added). This has been nothing if not an assessment of our methods, but perhaps Maddy's project is also not real philosophy. So, what would medical epistemology consist of if it hewed to a more restrictive conception?

I have already granted that the limits of inference, which I called "friction-free philosophy" in Chapter 2, can be applied without regard to social realities. However, because such questions ask about the warrant of evidence in best-case scenarios, they tell us very little about how to regard evidence where actual practice is far removed from the ideal. In reality, commercial imperatives are affecting not only the circulation of information, but the very constitution of evidence itself. Thus, the evaluation of warrant in any real-world scenario depends on an appreciation of the effect that commercial imperatives are having in that domain. Moreover, as we have already seen, altering regulations in an area that is ostensibly unrelated to epistemology (i.e., prohibiting advertising of off-label uses by the

FDA) can directly affect the integrity and reliability of a domain of canonical interest (e.g., the reliability of clinical trials for Vioxx). Accordingly, if EBM is to be a place in which philosophers are to have practical impact, then a more expansive framework is required.

7.2 Medical knowledge in a social world

In contrast to an epistemological focus on an isolated knower confronting a fixed set of evidence, social epistemology attempts "to come to grips with the social interactions that both brighten and threaten the prospects for knowledge" (Goldman, 1999, p. vii). Goldman's account provides a number of conceptual tools that help to situate my account of epistemic arms races, to which I add two further concepts: the notion of *reliability* and the notion of *robustness.*

7.2.1 The social epistemological framework

Goldman (1999) claimed that science is ultimately a social process, and one that cannot be understood apart from the social practices that play important roles in scientific discourse. In addition, he provided jargon and a framework to discuss the epistemological impact of interpersonal and institutional contexts that effect knowledge. Crucially, Goldman's social epistemology provides a way to discuss the beliefs of a social group. He coined the term *mental infosphere* to refer to all of the beliefs held by a relevant group, and defined human knowledge as the subset of beliefs that are true. Because some knowledge may be known by only a few, human knowledge can be increased by veritistically successful communication. Some communication is stored, so that the communicator need not be present when the message is received. Goldman referred to this collection as the *message infosphere*. Veritistic social epistemology aims to assess practices insofar as they affect the group's veritistic value (V-value), which

roughly equates to the extent to which the mental infosphere is populated with true beliefs on questions of interest. Veritistic analysis is concerned with how practices impact a group's V-value.[99]

As Goldman (1999) noted, not every actor needs to know all the same truths; rather, "institutions, for example, sometimes have an interest in certain of their actors knowing certain questions. It is appropriate for veritistic analysis to assess how well the institution's practices serve this end" (p. 94). Though Goldman struggled to specify how interest should factor into social epistemology generally, I would suggest that patient outcome provides a natural metric for medicine.[100]

I consider that it is not terribly important what a pediatrician believes about treating Alzheimer's, but it is quite important for a cardiologist to have true beliefs about heart arrhythmias. I also take it that the reason for this is that if the cardiologist has false beliefs about heart arrhythmias, their patient's health is at risk. Moreover, medical economics has developed measures such as the QALY (i.e., quality-adjusted life year), which could be used to further specify institutional interests. Setting aside technical debates about whether QALY is currently measured adequately, it is the sort of measure that would give very high priority to cardiologists knowing the truth about heart functions, very low value to pediatricians knowing truths about degenerative diseases that affect the elderly, and zero value to any doctor knowing the antepenultimate listing in the 1974 Omaha White Pages.

The final strength of Goldman's social epistemology is that it makes room for third parties. In addition to senders and receivers of messages, there are other agents that impact the flow of communication. In this class, Goldman includes journal editors, broadcasting sponsors, and government

---

[99] For a full defense of veritistic value see pp. 87–99; for Goldman's (1999) development of infospheres see Chapter 6, esp. pp. 161–165.

[100] There are some instances in which institutional veritistic interests and improvement in patient outcomes may not align (e.g., receiving a placebo in a trial), but such situations are uncommon. One large class of exceptions might be correct knowledge of an untreatable illness. For example, genetic testing can identify and reliably predict the age of onset for Huntington's chorea, though some elect not to be tested as there is no treatment. However, in most cases (e.g., terminal cancer) correct diagnosis should lead to proper palliative care and the avoidance of fruitless (and often harmful) interventions.

agencies such as the FCC and the FDA; essentially any party that alters the substance or changes the number of parties the message reaches. Even when these parties are not sending a message themselves, their actions have veritistic consequences by serving as "gatekeepers" of communication channels,

> given the importance of gatekeepers … social epistemology must inquire into the practices available to gatekeepers and the veritistic consequences that might flow from these practices. Casting our net more widely, we should examine not only the practices of individual gatekeepers, but the fundamental institutional arrangements or frameworks that influence the dissemination of thought and idea. (p. 189)

As examples of gatekeeping functions, Goldman discussed both explicit forms, such as government regulation against cigarette advertising, and indirect forms, such as when television companies refrain from airing a negative story about one of their advertisers.

7.2.2 Pharmageddon redux

With Goldman's framework, the concerns raised by medical ethicists in Chapter 1 can now be incorporated into medical epistemology. Presumably, it is now easy to see how seeding trials, manipulation of trial design, commercially funded medical education, and the financial pressures of the publishing industry affect the beliefs that doctors hold. For example, if a doctor begins to prescribe a drug inappropriately because they have been paid to serve as a key opinion leader, this will reduce the medical community's V-score. Though the routes are less obvious, problems such as disease-mongering and me-too drugs can lower the V-score as well.

Recall that in the 1970s osteoporosis simply referred to rapid loss of bone density, but has now been broadened so significantly that half of women will meet the criteria for it by age 52. The new guidelines emerged a year after a drug that increased bone density arrived on the market, and were put forward by a World Health Organization study group that had been sponsored by three pharmaceutical

manufacturers with products to treat osteoporosis (Abramson, 2004). This change was massively profitable, but it is not clear that the result has benefitted women's health. Critics (e.g., Abramson, 2004; Moynihan & Cassels, 2005) have alleged that by pathologizing a natural process, doctors prescribe drugs to the exclusion of addressing factors that are of equal or greater importance to preventing broken bones than bone density (e.g., smoking, muscle strength, suitable eye glasses, etc.). Let us assume that such critiques are well founded. Suppose it is the case that making lifestyle changes is better for preventing broken bones than pharmaceutical treatment, and that thinking about low bone density as a risk factor inclines doctors to recommend lifestyle changes whereas pathologizing it leads to pharmaceutical intervention.

Without getting into the debate about the objectivity of nosology, we can examine how such issues can be fit into the social epistemological framework. Assume there are some objective facts both about whether osteoporosis is a disease and what the best treatment is.[101] If osteoporosis is not a disease, then clearly the V-score would increase if doctors began seeing low bone density as a risk factor. In such cases, doctors acquire true beliefs and make better treatment decisions as a result. The interest that the medical community has in promoting this fact would depend on how much less risk women would face as a result of lifestyle modifications. On the other hand, if osteoporosis is a disease, but knowing this leads doctors to make worse treatment decisions, then the V-score would not increase until doctors also had true beliefs about how best to treat it.

A second problem that seem removed from epistemology was "me-too" drugs; that is, drugs such as Nexium that offer no substantial benefit over the drugs they replace (in this case Prilosec), but are marketed as breakthrough drugs. Though not clearly an epistemic problem at first blush, it is one of the central issues that faces medical social epistemology. As Goldman pointed out, the mere availability

[101] While there are not objective facts about nosology, there are still facts about best treatment.

of information is a prerequisite, but a superior communication system must also "ensure that interested readers locate and recognize intellectual products or documents that are evidentially appropriate to their projects and inquiries" (p. 175). The fact that the market is constantly flooded with new drugs that are promoted with equal enthusiasm, but where so few actually represent improvements, means that doctors simply do not have enough time to stay current with new products and separate the wheat from the chaff.

Though such products complicate the epistemic situation, the problem has not persisted for decades by mistake. As noted by Dr. Console, former head of research at E. J. Squibb, most me-too drugs were not pursued because companies thought they might provide a clinical benefit as "it is clear while they are on the drawing board that they promise no utility, they promise sales" (as cited in Mintz, 1965, p. 169; cf. Goodman, 1964). Thus, in both me-too drugs and disease-mongering we have practices that are massively profitable, but epistemically deleterious.

7.2.3 A pragmatic defense of social epistemology: How I learned to stop worrying and spot the bomb

The EBM movement has argued for its own conception of epistemology by asserting that practice would be improved if doctors were trained to base their treatments on evidence. Likewise, there is some reason to hope that if medical education provided a greater focus on social epistemology and doctors gained a broader understanding of what, in fact, affects their beliefs, they would be more circumspect about interactions with industry. There is at least significant evidence that doctors are unaware of how effective industry is at influencing them.

In an early study on the topic, Jerry Avorn asked doctors to indicate what shaped their prescription decisions (Avorn, Chen, & Hartley, 1982). Unsurprisingly, they indicated that research was

high on the list and advertising was largely irrelevant (these aspects were rated as very important by 62% and 3% of doctors, respectively). In the second part of the survey, Avorn asked doctors questions regarding therapeutic claims that were being heavily promoted, but which research had definitively shown to be untrue. A total of 71% of doctors cited the commercially promoted falsehood.

Orlowski and Wateska (1992) examined the effect on doctors at the Cleveland clinic attending a conference about two drugs in use at the time. Attendees were treated to an all-expenses-paid trip to attend the conference at a tropical resort. Afterwards, 20 doctors were interviewed and asked whether the conference had changed their opinions of the drug, to which 17 of the 20 reported that it had not. Despite the fact that, nationwide, there was no change in sales in the 22 months prior to the conference compared with the 17 months afterwards, usage of the first drug at the Cleveland clinic tripled and usage of the second drug more than doubled.

Finally, consider the more recent practice of hiring doctors to speak about pharmaceutical products. For example, Dr. Dickie was contacted by a detailer and agreed to become a paid speaker on behalf of a drug for attention-deficit hyperactivity disorder. As a result of publically avowing the benefits of the drug he became more convinced that it was the best available treatment on the market, and began prescribing it more frequently and in a wider variety of circumstances. He did all of this believing that he was educating other physicians, and that his prescription habits remained unchanged. When he was informed by an investigator that influencing *his* prescriptions may have been the real purpose of hiring him, he was disturbed, “because I perceive myself as always prescribing in the best interest of my patient. And even unconsciously, if I was unduly influenced, that would really bother me. I usually pride myself on keeping up my guard to prevent undue influence” (as quoted in Spiegle, 2010).

The central problem here is that as doctors have become more sophisticated consumers of research, pharmaceutical manufacturers have responded by presenting advertising in innocuous forms.

In the face of manufacturer tactics that are constantly evolving, what might be done to help doctors spot epistemic threats? Recall that in Chapter 3, soldiers faced a similar problem. As they became better at spotting IEDs, insurgents developed countermeasures such as placing IEDs inside roadkill or constructing false curbs. A major component of the American effort to defeat IEDs became intense training of the troops with the latest intelligence:

> we lowered the effectiveness of a single statistical IED in Iraq by a factor of six, which is not a trivial result… if you were to go to the national training center in California where the bulk of the army units train, you would find that the techniques used by the Iraqi American enemy [i.e., the simulated training enemy] change every six months to fit exactly what's going on down range [i.e., the battlefield] so every unit that comes through there is seeing the latest possible version of what they are going to expect downrange. And that takes a huge investment, a lot of care, a lot of feedback from downrange. However, it's worth it because when folks get there they have a leg up on the problem. (Gen. Meigs, 2010)

The problem with an epistemology that only focuses on trial design is that it sweeps under the rug many of the ways in which pharmaceutical manufacturers influence doctors. If EBM does not take into account this broader frame, guidelines for assessing evidence will continue to leave doctors susceptible to the effects of disease-mongering, disproportionate publishing, and the commercialization of medical trials. They will also fail to prepare doctors to negotiate the intense marketing pressure that is brought upon them by the competition created by me-too drugs. Ignoring the problem fails to prepare medical students for what they will experience in the field.

The strength of the social epistemological framework is that it makes a home for the various influences on prescription habits. Combined with a careful study of marketing practices, medical education could prepare doctors for the ways in which manufacturers will attempt to influence their prescription habits and train them to resist or avoid situations that would negatively affect the health of their patients. That said, doctors would still need to learn how to distinguish reliable evidence—an issue to which I now turn.

7.3 Disambiguating methodological rigor and epistemic reliability

In Chapters 3 and 4 I laid out the defining features of asymmetric arms races and showed that the history of medicine can be seen as an asymmetric epistemic arms race between "reformers," who attempt to improve medical care, and profit-driven pharmaceutical manufacturers. The arms race is asymmetric because the strategies employed by one side are not open to the other, and thus the evolution of strategies often takes the form of measure–countermeasure. In summary, the five essential features of asymmetric arms races are: (1) the effectiveness of any action typically decreases over time; this is because both (2) opponent's responses often attenuate the efficacy of one's actions and (3) opponents engage in a search process to identify and exploit weaknesses; however, (4) because actions are costly it is often disadvantageous to adopt new strategies until they are necessitated by an opponent's action; and (5) the process results in the gradual accumulation of costly strategies.

In the context of medical epistemology, the primary strategies of reformers have been to improve and implement research methodology and restrict advertising methods. The primary strategies of pharmaceutical manufacturers have been innovative forms of marketing, attempts to block the implementation of methodological rigor where possible, and attempts to capture and subvert influential levers of power if reform is implemented. In short, my account acknowledges that while there are forces that drive us towards pharmageddon, there have also been small groups of dedicated reformers pushing back.

A crucial desideratum for action is typically *reliability*. Something is reliable to the extent that it yields the desired outcome. Recall the asymmetric arms race between US soldiers and Iraqi insurgents.

To simplify, suppose that the only question we ask is, are high-powered radio jammers a reliable means of defending against IEDs? Such a question has no definitive answer. In April of 2005, when over 95% of IEDs were detonated by remote, the use of jamming technology was a reliable means of defense. By April of 2007, when most IEDs were detonated by pressure plates and command wire, jamming technology was far less reliable. The effectiveness of jamming technology decreased as it became prevalent and insurgents adopted other detonation mechanisms in response.

By this standard, it is not obviously true that RCTs are currently the most reliable form of evidence. Moreover, the longer RCTs are around, the more time manufacturers have to discover ways to manipulate the results. Safer (2002) compiled a number of means by which results of RCTs can be manipulated in ways that are extremely difficult, if not impossible, to detect.[102] While proponents of EBM promote RCTs on the basis that they are better able to control bias, we saw in Chapter 3 how the disproportionate weight the medical community now places on RCTs has made them an increasing target of manipulation. What is crucial to emphasize is that if such trials are scrutinized in terms of standard evidential hierarchies, industry-funded trials are *more rigorous* than independently funded trials (Lexchin et al., 2003). As a result, even granting the claim that RCTs are the most rigorous trial design, they may not be the most reliable form of evidence. I will examine two brief cases in which evidence from more rigorously designed trials was less reliable than that from less rigorous ones.

[102] Another example of subtle manipulation was relayed to me by an anonymous researcher involved in FDA trials for antidepressants. In order to minimize the appearance of a known (to company researchers) interference with cognitive functioning, the researcher scheduled patient examinations at a time of day when he knew the drug's negative effects would be at its daily minimum. He also chose a measure of cognitive function that would be insensitive to the types of cognitive deficits he expected the drug to cause. The entire protocol was negotiated with the FDA, but the researcher simply knew more about the drug than the official in charge of review. The data was honestly analyzed and reported, the methodology of the trial was completely rigorous (double-blind, randomized, etc.), but the results are hardly reliable evidence.

7.3.1 Broken codes are worse than worthless

In an asymmetric arms race, strengths have the potential to turn into liabilities as the other side first recognizes the strategy being employed and then seeks ways to circumvent it. A particularly apt example comes from the asymmetric contest between German U-boats and American anti-submarine technology (Meigs, 1990, pp.1 93–195). The German *Befehlshaber der Unterseebooten* (BdU) used communications intelligence to locate American convoys and direct the movement of U-boats at sea. A U-boat using BdU intelligence had a 250% greater chance of sinking a ship than a U-boat out on routine patrol. However, once Americans broke German codes, not only were German attacks six times less effective, the efficacy of American anti-submarine measures dramatically increased. A submarine on patrol had a 0.6% chance of being destroyed on a typical day, but once the code was broken a U-boat that received compromised BdU communication had nearly a three-fold increased risk of being sunk. Despite the shift in fortunes, the Germans never seem to have considered the fact that their code had been broken.

Here we have a case in which a reliable practice (sending coded messages) became a liability because its use continued after the underlying reason for its reliability no longer endured. If manufacturers have found ways to circumvent the methodological strictures of RCTs, then more methodologically rigorous evidence may in some cases be less reliable. Continuing to vaunt RCTs would then be unwarranted and in fact may be systematically misleading. I will argue that this has in fact occurred by showing that there are cases in which clinical experience, the least rigorous of all evidential sources, might trump RCTs. Even when clinical experience is formalized in a case series, evidential hierarchies rate such evidence below RCTs, cohort studies, and case-control studies (Howick, 2012; Strauss et al., 2011). I take it that if a case can be made that sometimes clinical experience should trump RCTs then, *a fortiori,* "stronger" evidence such as cohort studies may also do so.

Suppose you were a psychiatrist working during the late 1980s when Prozac first burst onto the market to widespread acclaim as a breakthrough drug. Amongst the claims made for Prozac is that it has a low incidence of side effects, in particular that the incidence of sexual dysfunction (impotence, decreased libido, inability to orgasm, etc.) is roughly 2–5%. Yet in your experience, sexual dysfunction has occurred in six out of the first 20 patients to whom you prescribe the drug. The odds that this has occurred by chance are already below 0.1%. Now of course there are several possible explanations for it, only some of which you can rule out. If you worked with a special population such as teenagers or the elderly this is a potential explanation for the discrepancy, but one that you could address by contacting colleagues. Nevertheless, the fact that all four published studies were funded by the manufacturer may reasonably loom large. If upon further discussions you find another five of your patients experiencing sexual difficulties that they had not brought up initially, you may be even more confident in judging your clinical experience as more reliable than the four industry-funded RCTs.

This suggestion does not absolve the doctor of responsibility from trying to rule out alternative explanations for the discrepancy, but neither are doctors afforded the luxury of waiting to act while further studies are completed. Given that manufacturers have a long history of understating risks and overstating benefits, and that doctors are generally biased in the same way, it is unreasonable to assert that doctors should treat the RCTs as a more *reliable* source of evidence just because RCTs are a more *rigorous* form of evidence. Insisting that the lone doctor discover the reason for the discrepancy may well be imposing an impossible burden on the physician. In this case, the reason why initial incidences were so low was that Prozac's trial protocol did not specifically query sexual dysfunction and so only patients that spontaneously mentioned their (potentially embarrassing) sexual problems to an independent rater were recorded. The fact that some manipulation of the results occurred is plain, though how such manipulation occurred is not. It is one of a number of cases in which something

became clinical wisdom long before being shown by more rigorous methods (in this case the actual rate of sexual dysfunction was approximately 60% (Montejo-González et al., 2000)).

It might be pointed out that among the reasons that clinical experience was reliable here was the large discrepancy between the reported and actual quantitative facts. However, it should be recalled from the previous chapter that in the case of antiarrhythmic-induced fatalities, a number of cardiac specialists noted that there were qualitative differences in many of the deaths (Velebit et al., 1982; Winkle, 1981). For example, Winkle noted that patients were exceptionally difficult to resuscitate. Similarly, qualitative differences were also noted in suicides while patients were on SSRIs (Healy, 2004).

While I am generally pessimistic about the reliability of clinical experience,[103] it may be too soon to dismiss clinical experience altogether. Suitably organized, clinical experience may be a valuable check on certain claims (e.g., frequency) even if it is unreliable in other domains (e.g., relative superiority). Furthermore, these examples merely serve to support the claim that there are circumstances in which less rigorous forms of evidence are nevertheless more reliable. I have far more confidence that this is applicable to the case of independently funded cohort studies vs. industry-funded RCTs.

7.3.2 The path of most resistance

One of the hallmarks of an asymmetric arms race is adaptation to the methods employed by the other side. The realization that consistency in behavior can ultimately be dangerous was key to training American soldiers and minimizing the threat posed by IEDs:

> I call this an arms race, where instead of taking years or decades to produce a new ICBM or a new radar it is weeks or days. They watch what we do … If they engage us with small arms, they know we kick out a machine gun team to the flanks to engage them behind a berm, that's where the IEDs are. They watch how we enter into courtyards. So the mantra for our soldiers and marines is "Take the path of most resistance." Slogging through the stream, climbing over great mounds, climbing over walls rather than going through openings in the

[103] For some horrifying cases of doctors torturing patients in the name of good health, see Whitaker (2002).

> courtyard. It's a thinking, smart, adaptive enemy we're fighting. (Lt. Gen. Barbero, 2012)

Just as IED emplacers attempt to plant IEDs where soldiers are likely to walk, pharmaceutical manufacturers place favorable evidence where doctors are most likely to learn.

It is a constant refrain that doctors do not have enough time to keep up with all of the treatments hitting the market each year. When asked why doctors see detailers, the most common answer cited is facility and efficiency:

> I think the answer is it's user friendly, it's very user friendly and it's easy listening, you know, with your coffee listening to what they've got to say. I'm sure you could manage if you didn't see another drug rep and I'm sure you could get the information if you wanted to, it's just that it's not that accessible, and it's also whether you would actually have the time to sit and read it. (as quoted in Prosser & Walley, 2003, p. 307)

Detailers, industry-funded continuing medical education, medical conference exhibits passing out reprints, and even advertising, all present easy paths for doctors to gather evidence about new treatments. The need to stay up to date on the proliferation of new treatments (exacerbated by the profusion of me-too drugs) combined with the effort it would take to do so independently creates conditions that are ripe for commercial exploitation.

Furthermore, even supposing that a doctor adopted the policy of staying up to date with the research literature, this effort alone will often be insufficient. Suppose that at the end of the year in 1986 a doctor decided to identify and read every article published on class 1 antiarrhythmic drugs published in the past year in either *Circulation* or *The American Journal of Cardiology.* We know from Chapter 2 that at the time there was good reason not to widely prescribe such drugs, yet such a search would identify nine studies, every one of which was positive.[104]

[104] This is based on a manual search of the two journals looking at articles that mentioned arrhythmias, VEBs, or any antiarrhythmic drug. This figure does not include any of the many articles discussing the usefulness of research methods such as electrical stimulation or Holter monitors as a means of assessing treatments. It also does not include single-case studies or editorials. Many studies identified risks of the antiarrhythmic drug, but

The epistemic consequences are two-fold. First, doctors must take the path of most resistance if they wish to locate unfavorable information. For various reasons, critical information will often be published in lower-tier journals, if it is published at all. Secondly, because unfavorable studies are less likely to be published and because of the professional costs that are frequently associated with choosing to publish unfavorable results, such evidence should be accorded disproportionate evidentiary weight. As we saw in Chapter 2 with the blacklisting of Winkle after publishing the negative results he obtained with Enkaid, researchers take great personal risk by publishing unfavorable findings, and such results should thus be treated as more reliable.[105]

Lastly, identifying reliable sources (either researchers or publications) may be one way that doctors can make the path of greatest resistance less onerous. For example, if doctors had ignored the articles discussed above and consulted *Medical Letters*,[106] they would have been instructed to only use Tambocor for "serious ventricle arrhythmias that have not responded to older, better established drugs" (*Medical Letters*, 1986, p. 20) and known that "whether any drug should be used to treat asymptomatic or mildly symptomatic conditions remains to be determined … [as drugs] can aggravate arrhythmias" (*Medical Letters*, 1987, p. 52). Abiding by the advice of *Medical Letters* would require radically devaluing the results of most published research and the opinion of many leading experts (particularly those with industry ties). It would also have saved many patients' lives.

---

every study concluded something to the effect of "This drug is clinically efficacious and safe." A study would have been judged as negative if it had included a statement such as "This drug was not effective and/or was unsafe to administer." The only negative article I discovered was an editorial calling on doctors not to treat asymptotic cases until further studies were done (Josephson, 1986). In light of the CAST study published three years later, such a position seems prophetic.

[105] The disproportionate weight accorded to negative evaluations is not without limits. For example, critiques of psychiatry are not against the author's interests if they are a member of the Church of Scientology, which regularly protests all psychiatric medications. Similarly, when the negative appraisal of a drug is funded by a drug's competitor, the disproportionate weight principle clearly does not apply. Sadly, such mechanisms are often how unfavorable facts become established in the literature, as when the manufacturers of Wellbutrin began funding studies on the sexual side-effects of Prozac (e.g., Segraves, Segraves, & Bubna, 1995; Walker et al., 1995).

[106] *Medical Letters* has a long history of independence, bans on conflicts of interests, and carries no advertising.

Though it is rare, official bodies have occasionally considered devaluing the epistemic weight accorded to industry-funded studies. In the wake of the realization that manufacturers had withheld negative studies on SSRI suicide data, the British advisory agency NICE considered, but ultimately rejected, such a policy. The present analysis suggests that this decision could have been justified by averting to the reliability standard.

7.4 Recognizing constraints on interactions and robustness

If the argument thus far is correct, then our current epistemic practices do not track reliable evidence and must change. A common method of generating solutions is to examine current and past threats and to devise countermeasures or new practices. In this chapter, I will argue that while such efforts are necessary, they are not sufficient, and I will then introduce a standard that solutions should be judged by. Perhaps the first insight that falls out of seeing an interaction as an arms *race*, is that thinking in terms of solutions is misleading; it is better to think in terms of an ongoing contest. In judging possible responses there are two distinct lessons from other asymmetric arms races. The first is that sometimes the most visible aspect of the problem is not the most effective place to attempt to intervene. Second, not every effective intervention is robust, and it is robustness that is crucial in an arms race.

7.4.1 Misleading appearances: The importance of recognizing constraints

To observers of brood parasitism one of the strangest occurrences is the absence of nestling rejection. Given that parasitized species will reject eggs on the basis of subtle cues (eggs laid at the wrong time, that have the wrong color or markings, are too large, etc.), why do reed warblers fail to realize that the chick they are feeding is now larger than they are? This is not a terribly subtle cue and is one a bird could reasonably be expected to notice. Dawkins and Krebs (1979) supposed

that the cry of the cuckoo was somehow manipulating their reed warbler foster parents:

> Those who have never been brainwashed or addicted to a drug find it hard to understand their fellow men who are driven by such compulsions. In the same naïve way we cannot understand a host bird's being compelled to feed an absurdly oversized cuckoo … With natural selection working on the problem, who would be so presumptuous as to guess what feats of mind control might not be achieved? Do not expect to see animals always behaving in such a way as to maximize their own inclusive fitness. Losers in an arms race may behave in some very odd ways indeed. (Dawkins & Krebs, 1979, p. 64)

Dawkins and Krebs suggested that the host is controlled by the chick's siren song. How else might one explain such keen discrimination over unresponsive eggs and such wanton disregard once the parasite can begin signaling? Since feeding the cuckoo is clearly against the interests of the reed warbler, the cuckoo must find a way to exercise control and ensure that the reed warbler disregards its own best interests.

The same appearance of control and lack of resolve has captured the focus of medical reformers. As discussed in Chapter 4, early reformers initially believed that the prescription of worthless products was due simply to absence of an independent sources of information to facilitate better judgments. Yet after years of exposing fraud with little change in doctors' prescription habits, most of the council's venom turned from manufacturers to the doctors who could no longer claim ignorance as an excuse. The hope that doctors would awaken to their responsibility, cast off the control of pharmaceutical companies, and raise the standards of medicine was eloquently expressed by Robert Hatcher:

> Ours is considered a learned profession; we enjoy great privileges and have important duties to the public; we hear it asserted frequently that our profession is engaged in a great struggle to alleviate suffering and to ameliorate the condition of mankind. It is indisputably true that a vast amount of our energy is thus directed, but unfortunately, it is just as indisputably true that we have permitted ourselves to become "dupes and accessories" of nostrum vendors; we have continued to prescribe nostrums that have proved to be fraudulent … we use drugs recklessly when we are employed to save life and are trusted as men in no other profession; we take words out of the mouth of ignorance and proclaim them with authority; we take statements without evidence to support them and on them base the treatment of the sick; I would that I had the power to carry

> home to every individual member of the medical profession a sense of his duties to the public, to his profession and to himself so that he might see himself in his true light, for I believe that our failure to correct these abuses is, as stated in the report which I have frequently quoted, due to inertia and not to a deliberate neglect of a duty. (1916, p. 1,314)

Hatcher felt that doctors had fallen under the spell of pharmaceutical manufacturers by convincing themselves that they personally were not unduly influenced by industry largess. However, unlike pharmaceutical manufacturers (who have a fiduciary responsibility to their stakeholders) and patients (who typically lack the ability to read and digest the scientific evidence), doctors occupy a unique position and have correspondingly unique duties to remain financially and intellectually independent. Indeed, a modern rendition of Hatcher's argument was offered by medical ethicist Howard Brody (2007) in his aptly titled *Hooked.* He argued that doctors have grown dependent on the perquisites of the pharmaceutical industry, and as a result are serving the interests of manufacturers instead of patients.

While it is certainly true that industry has contributed to irrational prescription habits, the addiction metaphor is inadequate. Returning to the reed warbler, more recent research has shown that nothing about the call of the cuckoo chick leads to the absence of nestling rejection. It is simply that given the constraints of the interaction, the cost of occasionally raising a cuckoo outweighs the cost of evolving nestling rejection (Lotem, 1993).

The reason for this is that birds learn via imprinting. If a reed warbler imprints on her first clutch and it is parasitized she will imprint on both her eggs and cuckoo eggs. This will impair her ability to identify and reject cuckoo eggs in the future, but on occasions where she is not parasitized all will go smoothly. In contrast, because a cuckoo chick's first instinct is to destroy all the other eggs in their nest, the cuckoo will be the sole nestling. This combination of learning by imprinting and the cuckoo chick's destructive behavior means that the reed warbler faces extreme costs if she rejects odd-looking nestlings. If her first clutch is parasitized, she will imprint on the cuckoo chick and never raise a chick of

her own. In contrast, in species where the brood parasite is raised alongside nestlings from the host species there is both nestling rejection by the host and nestling mimicry by the parasite (Fraga, 1998). While there are clear cases of manipulation by parasitic chicks (e.g., supernormal gape markings), manipulation is only part of the story. In this case, what appears to be manipulation is in fact rational behavior on the part of the reed warbler, given the constraints of nature (Davies, 2000).

On the basis of manufacturer-induced irrational prescription habits, many authors have been quick to suggest that doctors are being manipulated (Abramson, 2007; Angell, 2004; Avorn, 2004; Brody, 2007; Kassier, 2005). While there are clear cases of manipulation by manufacturers, manipulation alone cannot explain a century of reliance on manufacturers despite evidence of unreliability. While I agree with the sentiment, top-down regulation and moral suasion alone are insufficient. The fundamental question of why no such pushback exists in the first place is a good indication that doctors are getting far more out of their relationship with pharmaceutical manufacturers than a supplemental income.

The first criterion of a good response is that it is based on an appreciation of the constraints on action. There is a great need for a full and convincing account of why doctors are so heavily tied to industry, and it will not do to excuse their behavior with ignorance or irrationality. Merely trying to block the relationship between doctors and manufacturers does little to address the underlying forces that are driving doctors into relationships with manufacturers in the first place. Recommendations that fail to take into account such dynamics will be as effective as earplugs would be for reed warblers.

7.4.2. Red teaming and robustness: Establishing a standard for future innovations

The biggest difference between the types of biases I identified in Chapter 2 (error and manipulation) is the standard that is applicable in judging which possible policy changes are likely to be successful. To

deal with errors, a solution must be reliable; to deal with manipulation, a solution must be robust. The difference can be easily seen by considering the electromagnetic predetonator.

An IED is essentially a power source, a trigger, and a charge. The trigger connects the power source to the detonator and sets off the main charge. One of the early counter-IED (c-IED) innovations was a device that sent an electromagnetic pulse ahead of the vehicle to destroy the internal circuitry of the trigger. The pulse caused the trigger to either "fail closed," in which case the IED exploded before the vehicle was within range, or "fail open," in which case the IED would no longer detonate. Either way, the predetonator was a highly effective piece of c-IED technology. It was also a colossal waste of money. Almost as soon as the technology reached the battle field, insurgents devised a means of negating the effectiveness of the predetonator. As noted by electrical warfare specialist Daniel Widdis, it would have been better to do nothing than to field "billion-dollar solutions with ten-cent countermeasures" (as cited in Zorpette, 2008). The predetonator was both extremely effective and extremely fragile, and so a complete failure. This case illustrates why robustness is the proper criterion of a good response.

In order to avoid such disastrous projects the pentagon has begun "red teaming," in which a select group of intelligence professionals and explosive experts are tasked with designing counter-measures to American innovations. Such activities can be useful in both screening out effective but fragile solutions, and anticipating and designing next-generation capabilities before they are necessitated. A crucial aspect of red teaming is first assessing whether any of the available strategies circumvent the proposed innovation. If the answer to this is negative, red teaming provides some insight as to how robust the solution is likely to be. An effective strategy that is circumvented by an already existing (but rarely employed) countermeasure will simply precipitate a rapid shift to the

countermeasure. An effective strategy that has an easy and obvious counter will be effective only so long as the other party does not realize the strategy is being employed.

Using the standard of robustness, it is plain that many solutions proposed by reformers are inadequate. For example, earlier I argued that discounting research funded by industry could be justified because of solid evidence that shows that industry-funded research is unreliable (e.g., Bekelman, Li, & Gross, 2003; Lexchin et al., 2003). However, suppose that such a standard was employed and became decisive in medical decisions. The natural response from industry would be to conceal funding. Surely anyone who lived though the 2012 presidential election recalls that the consequence of requiring funding disclosures for political ads was the explosion of political organizations created to obscure the source of funding. A move to devalue the evidential import of commercially funded research is extremely fragile without a robust means of tracking finances.

The same manner of critique can rule out a number of well-meaning suggestions. For example, Howick (2012) suggested that placebo-controlled trials should be replaced with trials that compare the new treatment with the standard one, the guiding idea being that what we would like to know is that new treatments represent substantial improvements on available care, not just that treatments are better than placebos. However, given the considerable evidence that manufacturers can manipulate comparison trials just as easily as placebo-controlled trials (Bero et al., 2007; Rochon et al., 1994; Safer, 2002), such a recommendation fails because it is susceptible to currently available strategies. As a result, a move towards comparison trials in and of itself is unlikely to lead to substantially improved treatment options.

An example of a fragile solution is Howick's (2012) suggestion that blinding statisticians will prevent analysts from producing favorable analyses: "If data analysts did not know which was the experimental therapy, they would not be able to predictably make systematic errors or omissions that

favored the apparent benefit of the experimental drug" (p. 191). However, simply giving the statistician a "0" and a "1" rather than "treatment" and "placebo" will not prevent such abuses. Any statistician worthy of being hired will be able to distinguish the two groups, even if they are unlabeled. For example, a cursory examination of the side-effect profiles of two groups will immediately reveal which group is which. The upshot is not that such issues should not be addressed; rather, it is that care must be exercised before recommendations are enacted. Masking statisticians in conjunction with other measures may be reasonable; however, the strategy without further elaboration is extremely fragile.

The preceding argument has some puzzling consequences. An epistemic strategy may be reliable in the sense that at a given time using such a method will provide the most accurate estimate of treatment efficacy, yet instituting the strategy would be inadvisable because of the fragility of the response. To differentiate the two, we might distinguish between synchronic and diachronic reliability. A strategy is synchronically reliable if it provides accurate estimates given the current available evidential practices. A strategy is diachronically reliable if the implementation of the strategy cannot be undermined by known or easily generated counterstrategies (viz., the solution is robust).

7.4.3 Systematic analysis

The benefit of Goldman's social epistemology is that it shifts the focus of medical epistemology from a disembodied inference practice to the factors that affect the beliefs practicing doctors actually hold. The pertinent question then becomes what set of practices (i.e. evidence gathering, dissemination, and amalgamation) should a community adopt in order to maximize the truth of practitioners' beliefs? To this I add that in arms races the answer to this question must be judged in terms of the robustness of proposed strategies.

As an example, recall that Howick framed his book as "an evaluation of the EBM view of what counts as 'good evidence'" (Howick, 2012, p. 24). Now, let us reconsider the question posed by Howick (2012): (1) Was the meta-analysis performed by Patricia Crowley (1981) sufficient evidence to make treatment of infants with corticosteroids standard practice, despite the fact that experts judged corticosteroids to provide no benefit (Howick, 2012, p. 161)? Notice that this question might be interpreted in many ways. The first is: (1a) After Crowley's meta-analysis, did the sum total of available evidence (including expert judgment) support the use of corticosteroids in premature births? This is essentially a question regarding what inference should have been drawn if someone looked at all the evidence available at the time.

A different reading of the question is: (1b) Given the communication structures and normative epistemological practices of the community of obstetricians, should the use of corticosteroids in premature births have become standard practice? As a matter of historical fact, Crowley's meta-analysis did not change practice even though it showed a definitive reduction of infant mortality. Nor did it change in 1989 when she strengthened her results with additional studies. Indeed, practice did not evolve until 1994, when the NIH conducted a study and published its consensus statement in *JAMA*. When framed in this way, a greater number of possible foci are available. Crowley's original study was published in a journal's first issue and the second article was published in an edited volume. Question (1b) allows that there might have been reasons for a failure to change other than doctors not knowing how to weigh evidence (e.g., deficiencies distributing evidence to practicing physicians).

Finally, there is a more systematic way to construe the question: (2) Would the medical community have a higher V-value if it instituted procedures such that all doctors were apprised of and accepted the results of studies like Crowley's? As discussed in the previous chapter, Zollman's (2007) results suggested that in some cases immediate communication of results to the whole community is a

detriment to achieving a true consensus. Moreover, we have seen that in the case of Vioxx, the result of doctors being trained to act on studies such as Crowley's were pharmaceutical companies publishing systematic reviews that falsely identified their products as superior to traditional NSAIDs. In light of such considerations, it is not obviously irrational to wait for an authoritative independent body like the NIH to recommend a practice before adopting it. From a policy perspective, one has to at least consider how pharmaceutical companies will react if such practices become widespread.

7.5 Envoy

Beyond what has been reviewed above, there may not be much more that can be said about the best way to analyze data in general. Once the tie between methodological rigor and reliability is severed, we are left without a generally superior method to evaluate evidence. In a field that is heavily dependent on industry research, the most reliable source of evidence may be observational studies. In an area that is dominated by independent researchers, RCTs may be preferred. Moreover, if observational studies were synchronically more reliable, it may still not be justified to change evidential standards. Worrall (2002, 2007a, 2007b, 2010) has argued that an observational study may provide just as good evidence as a randomized trial if one controls for known confounds. Supposing we accept this argument, should we adopt a policy of treating such trials as equivalent to RCTs?

On one hand, observational trials may be less likely than RCTs to be industry funded in a given field, so there may even be reason to judge such trials as more synchronically reliable. However, the clear impact of increasing the evidential weight accorded to observational studies would be an increase in the number of industry-funded observational studies. The ability to control for variables post hoc greatly expands the ability to manipulate the data, so while there may be a friction-free justification (i.e., a justification in an epistemically perfect world) for equating well-controlled observational studies

with RCTs, it may yet be a diachronically unreliable (i.e., fragile) strategy. On the other hand, since observational studies are cheaper to conduct it may open up an avenue for independent researchers to contribute to the knowledge base in a way that outweighs the increased abuses. Adjudicating such a question is beyond the scope of the present work, but it is the type of analysis that is absent in philosophic discussions and clearly mandated once medical epistemology is understood as an asymmetric arms race.

In closing, I might note that for some, there is something profoundly unsettling and indeed unsatisfying about this account: it lacks a definitive answer. At the end of a journey such as this, it is the author's job to provide a formula that allows everything to fit into place and gives the reader hope that the problems identified can be solved once and for all with the newfound insight. No such answer has been provided here. Perhaps this is a shortcoming, but it is, for better or for worse, the conclusion of the dissertation. If the account provided is correct, there is nothing more to offer.

Nevertheless, the impossibility of a decisive solution can also be freeing. It suggests that a new course is called for, one that does not ignore the veritistically disruptive effects of commerce. I propose that to make veritistic gains we have to embrace the suck: to accept that there are forces working against attaining reliable knowledge, to study their tactics, and develop countermeasures. Chapter 5 suggested the reform of guidelines for reporting and analyzing data. Chapter 6 suggested that the communication of information might be improved if we institute procedures for discounting unreliable information. Both of these countermeasures should be red teamed to determine whether they are robust, but they are plausible countermeasures. There are no doubt others.

As seen in Chapter 1, many despair at the state of medical practice. Such despair is not new. Voltaire is said to have lamented, "Doctors are men who pour drugs of which they know little into

patients about whom they know less with diseases of which they know nothing."[107] Oliver Wendel Holmes (1860), dean of Harvard's medical school, wrote: "I firmly believe that if the whole of *material medica* [i.e., drugs], as now used, could be sunk to the bottom of the sea, it would be all the better for mankind—and all the worse for the fishes" (as quoted in Parascandola, 1992, p. 8.). However, there is no justification for such therapeutic nihilism today.

Though my father might have died if he had taken Seroquel, the only reason he lived another five years was because of drugs that prevented his immune system from rejecting the lung transplant. That was five more years of love, good arguments, and fatherly advice. He got to meet my partner, reconcile with my brother, and make many delicious desserts. He got to read and discuss much of this dissertation. None of that would have happened without modern medicine. Profit-driven medical research can and does yield real advances. It will continue doing so to the extent that the market rewards genuine breakthroughs. The goal of reform should not be the elimination of private interest from medicine, but the continual corralling of such interests into epistemically reliable channels. We should not expect to be able to develop an incorruptible method; the goal of identifying the perfect experimental design or inference pattern is chimerical. There is no final resolution to the fundamental antagonism between commercial and veritistic forces. There is only a next move.

[107] The quote is also attributed to Molière, which means it was probably said by neither. It dates back to at least the early 1900s, where it is occasionally repeated in medical journals unsourced.

REFERENCES

Abraham, J. (2002). "The Pharmaceutical Industry as a Political Player," Lancet, 360: 1498–502.

Abramson, J. (2004). Overdosed America. New York: Harper Collins.

Adams, S. H. (1907). The Great American Fraud, Chicago: P.F. Collier and Son.

Adamson, W. (2007). An Asymmetric Threat Invokes Strategic Leader Initiative: The Joint Improvised Explosive Device Defeat Organization. Washington, D.C.: Industrial College of the Armed Forces.

Alexander, J.. (2013). "Preferential Attachment and the Search for Successful Theories". Philosophy of Science 80: 269-82.

Altman L. (1999). "Inside medical journals, a rising quest for profits." New York Times, August 24: F1.

Anderson, J., Fainaru, S., and Finer, J. (2005). "Bigger, Stronger Homemade Bombs Now to Blame for Half of U.S. Deaths," The Washington Post, October 26, p. A1.

Anderson, J., Lutz, J., Allison, S. (1983). "Electrophysiologic and antiarrhythmic effects of oral flecainide in patients with inducible ventricular tachycardia." Journal of the American College of Cardiology, 2, 105-114.

———, Stewart, J., and Crevey, B. (1984). "A proposal for the clinical use of Flecainide." American Journal of Cardiology, 53, 112B-119B.

Angell, M. (2004). The Truth About Drug Companies. Random House Press. New York.

Antonuccio D., Danton, W., and DeNelsky, G. (1995). "Psychotherapy versus medication for depression: challenging the conventional wisdom with data." Professional Psychology: Research and Practice, 26, 574–85.

Armstrong, J., (2001). Principles of Forecasting: A Handbook for Researchers and Practitioners. Norwell: Kluwer.

Atkinson, R. (2007). "Left of Boom: The Struggle to Defeat Roadside Bombs," Washington Post Special Report, September 30, A13.

Austin, A. E. and Larrabee, R. C. (1906). "Acetanilid Poisoning from the Use of Proprietary Headache Powders," JAMA, 46: 1680-1681.

Avorn, J. (2004). "Powerful Medicines: The Benefits, Risks, and Costs of Prescription Drugs." New York: Alfred A. Knopf.

———, Chen M, Hartley R. (1982). "Scientific versus commercial sources of influence on the prescribing behavior of physicians." American Journal of Medicine, 73, 4-8.

Ball, J. and L. Kiloh. (1959). "A Controlled Trial in Treatment of Depressive State." British Medical Journal, 2, 1052-1055.

Bamigboye, A., and Morris, J. (2003). Oestrogen supplementation, mainly diethylstilbestrol, for preventing miscarriages and other adverse pregnancy outcomes. Cochrane Database of Systematic Reviews 3: Art. No.: CD004353

Barbero, M. (2012). Countering the threat of IEDs. Presentation to the International Institute for Strategic Studies. February, 21

Bass, A. (2008) *Side Effects*. Chapel Hill, N.C.: Algonquin Books.

Bastedo, W.A. (1915). "Clinical Experience with Liquid Parafin (Liquid Petrolatum): A Comparative Investigation Made Under the Auspices of the Council on Pharmacy and Chemistry," *JAMA,* 64*:* 808-809.

Bekelman, J., Li, Y. and Gross, C. (2003). "Scope and impact of financial conflicts of interest in biomedical research." JAMA, 289: 454-65.

Bero, L., Oostvogel, F., Bacchetti, P. and Lee, K. (2007). "Factors associated with findings of published trials of drug-drug comparisons: why some statins appear more efficacious than others." *PLoS Medicine*, 4: e184.

Bhandari, M., Busse, J., Jackowski, D., et. al. (2004). "Association between industry funding and statistically significant pro-industry findings in medical and surgical randomized clinical trials." *Canadian Medical Association Journal*, 170: 477-480.

Bigger, J. Jr. (1984). "Antiarrhythmic Treatment: An Overview." *American Journal of Cardiology*, 53: 8B-18B.

————, Fleiss J., Kleiger R., et. al. (1984)." The relationship between ventricular arrhythmias, left ventricular dysfunction and mortality in the two years after myocardial infarction." *Circulation,* 69: 250-258.

Blakeslee S. (1998). "Placebos Prove so Powerful Even Experts Are Surprised." *New York Times (Science Times)*, Oct 13, 1998, p. C5.

Brody, H. (2008). *Hooked: Ethics, the Medical Profession, and the Pharmaceutical Industry*. Lanham, Md: Rowman and Littlefield.

Brooke, M. and Davies, N. (1988). "Egg mimicry by Cuckoos Cuculus canorous in Britain." *Journal of Animal Ecology*, 56: 873-883.

Brooker, M. and Brooker, L. (1991). "Eggshell strength in cuckoos and cowbirds." *Ibis*, 133: 406-413.

Broom, M., Ruxton, G. and Kilner, R. (2008). "Host life history strategies and the evolution of chick-killing by brood parasitic offspring." *Behavioral Ecology*, 19: 22-34.

Brownell, K. and Stunkard, A. (1982). "The double-blind in danger: Untoward consequences of informed consent." American Journal of Psychiatry, 139: 1487-1489.

Buck, H. C. (1909). "What is the Future of the Medical Profession?" *The New Charlotte Medical Journal,* 59*:* 28-30.

Bull, J.P. (1959). "The Historical Development of the Clinical Therapeutic Trial," *Journal of Chronic Diseases,* 10*:* 218-248.

Burrow, J. (1963). *AMA: Voice of American Medicine.* Baltimore: Johns Hopkins Press.

Bystritsky, A., and Waiker, S. (1994). "Inert placebo versus active medication: Patient blind ability in clinical pharmacological trials." Journal of Nervous and Mental Disorders, 182: 485-487.

Campbell, E., Gruen, R. Mountford, J. Miller, L., Cleary, P. and Blumenthal, D. (2007). "A national survey of physician-industry relationships." *New England Journal of Medicine*, 356: 1742-1750.

Carlat, D. (2010). *Unhinged: The Trouble with Psychiatry—A Doctor's Revelations about a Profession in Crisis*. New York: Free Press.

Carpenter, D. (2010). *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA.* Princeton: Princeton University Press.

Carter, A. and Gilmore, J. (2012). "Running the numbers on MRAPs: Reliable data proves the vehicles are worth the money." *Foreign Affairs*, <online archive>

Cartwright, N. (2007). "Are RCTs the gold standard?" *Biosocieties*, 2: 11-20.

———— (2009). "What is this Thing Called 'Efficacy'? in Philosophy of the Social Sciences." *Philosophical Theory and Scientific Practice.* (ed) C. Mantzavinos. Cambridge: Cambridge University Press.

———— (2010). "What are randomized controlled trials good for?" *Philosophical Studies*, 147: 59-70.

———— (2011). "A philosophers view of the long road from RCTS to effectiveness." *The Lancet*, 377: 1400-01.

———— and Hardie, J. (2012). Evidence-based Policy. Oxford: Oxford University Press.

Cary and Youssef, (2011). "JIEDDO: The Manhattan Project that bombed." The Center for Public Integrity <online archive>

CAST investigators. (1989). "Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction." *New England Journal of Medicine*, 321: 406-12.
CAST II investigators. (1992). "Effect of the antiarrhythmic agent Moricizine on survival after myocardial infarction." *New England Journal of Medicine*, 327: 227-233.
CenterWatch. (2002). "Phase IV Market Steams Ahead." CenterWatch, October.
Chisholm, P. (2005). "Clearing the roads." *Special Operations Technology*. <Online Archives: 3, v. 6>
Cho, M., and Bero, L. (1996). "The quality of drug studies in symposium proceedings." *Annals of Internal Medicine,* 124: 485–489
Colwell, N.P. (1909). "The Need, Methods and Value of Medical College Inspection," *JAMA,* 55: 513-514.
Cowley, A., Skene, A., Stainer, K., and Hampton, J. (1993). "The effect of Lorcainide on arrhythmias and survival in patients with acute myocardial infarction: an example of publication bias." International Journal of Cardiology, 40: 161-166.
Council on Pharmacy and Chemistry (CPC). (1905a). "Acetanilid Mixtures," JAMA, 44: 1791.
——— (1905b). "Sal-Codeia Bell," JAMA, 45: 1422.
——— (1906a)."Hagee's Cordial of Cod-liver Oil: Fraud and Deception Connected with So-Called Cod-Liver Oil Preparations," JA MA, 47: 1207-1208.
——— (1906b). "Phenalgin: A Typical Example," JAMA, 46: 290.
——— (1907a)."Labordine," JAMA, 48: 1121-1123.
——— (1907b). "Campho-Phenique," JAMA, 48: 1365.
——— (1907c). JAMA, 48: 1962.
——— (1908a)."Adding Insult to Injury," JAMA 50: 467.
——— (1908b). "Antikamnia and the Public: Samples, Form Letters and 'Prescriptions' Sent to the Laity," JAMA, 50: 1281.
——— (1908c). "Narkine: The Intangible Product of the Tilden Laboratory," JAMA, 51: 1439.
——— (1909). "Migranin," JAMA, 52: 1851-1852.
——— (1910a)."The Firm Replies," JAMA, 54: 1216.
——— (1910b). "Camphenol," JAMA, 55: 1662.
——— (1910c). "Aspiro-Lithine," JAMA 54: 1803.
——— (1910d). "Cactus Grandiflorus," JAMA 54: 889.
——— (1911a)."Formurol," JAMA, 56: 210.
——— (1911b). "Papine: A Disguised Morphin Solution," JAMA, 56: 1278.
——— (1912a)."An Ethical (?) Proprietary Exploited Under Fraudulent and Lying Claims," JAMA, 58: 280.
——— (1912b). "Tyree's Antiseptic Powder: Now Advertised Direct to the Public as the 'Best Preventative Known,'" JAMA, 59: 666
——— (1913a)." The Army and Navy Medical Record: A fraudulent publication whose editorial opinions are for sale," JAMA, 61: 1553-1555.
——— (1913b). Trypsogen, JAMA, 61: 1649.
——— (1913c). Sinkina, JAMA, 61: 1056.
——— (1914a). "Amourous Phosphorus: Dr. Nascher's Reply to The Journal's Article-Comments," JAMA, 62: 1033-1035.
——— (1914b). "Glyco-Thymoline," JAMA, 63: 1312-1313.
——— (1914c). "The Hypophosphite Fallacy: An example of the perpetuation of a false theory by advertising," JAMA, 62: 760-762

——— (1914d). “Buffalo Lithia Water: Contains One-fifth as much lithium as the Potomac River water,” JAMA, 62: 834-835.
——— (1914e). “Expurgo (Sanol) Anti-Diabetes: One More Fraudulent Nostrum for Diabetes,” JAMA, 62: 312-313.
——— (1914f). “Santogen: A scientific investigation of alleged action on recuperating powers of the blood." JAMA, 63: 1127-1129
——— (1914g). “Amorphous Phosphorus: A Practically Inert Substance Introduced as a Valuable Therapeutic Agent,” JAMA, 62: 79 3-794.
——— (1915a).“At the Bar of Public Opinion,” JAMA, 65: 1373.
——— (1915b). “Bell-ans (Pa-pay-ans Bells): Another 'Patent Medicine' Foisted on the Public Through the Medical Profession,” JAMA, 64: 1815-1816.
——— (1915c). “Tri-iodides, Three Chlorides and Maizo-Litium,” JAMA, 64: 528-530.
——— (1915d). "Neurilla," JAMA, 64: 1093.
——— (1915e). "Articles Refused Recognition (Venodine; Standard Radium Solution for Intravenous Use)," JAMA, 64: 2155.
——— (1915f). "Hydragogin," JAMA, 65: 894.
——— (1916a).“Whose bread I eat, his song I sing,” JAMA, 66: 38.
——— (1916b). “Mark White Goiter Serum and Mark White Iodinize Oil,” JAMA, 67: 969-970.
——— (1916c). “Why Glycerophosphates?” JAMA, 66: 1205-1206.
——— (1916d). “Intravenous Therapy,” JAMA, 67: 1450-1451.
——— (1917a).“Pil. Cascara Compound-Robins,” JAMA, 68: 303.
——— (1917b). “Firolyptol Plain and Firolyptol with Kreosote,” JAMA, 68: 564.
——— (1917c). “Biniodol,” JAMA, 68: 650.
——— (1918a).“Medical Society of the United States: From ‘Division of Fees’ to ‘Down with Autocracy,’” JAMA, 71: 1158-1159.
——— (1918b). Holadin and Bile Salt Mixtures, *Reports of the Council on Pharmacy and Chemistry*: 59.
——— (1918c). Shotgun Nostrums, JAMA, 70: 642-643.
——— (1918d). "Carminzym," JAMA, 71: 1081-1082.
——— (1918e). "Liquor Santaiva, S. and D., Omitted from N. N. R," *Reports of the Council on Pharmacy and Chemistry*, p. 66.
——— (1919a). "The Allied Medical Associations of America: Another Rocket in the Pyrotechnics of Quasimedical Organizations, " JAMA, 73: 52-54.
——— (1919b). “Commercial therapeutics,” JAMA, 73: 109.
——— (1920). "Cooperation of Pharmaceutical Houses." JAMA, 74: 1235.
——— (1920). Cooperation of Pharmaceutical Houses. JAMA, 74, 1235
Crowley, P. (1981). "Corticosteroids in pregnancy: the benefits outweigh the cost." *Journal of Obstetrics and Gynaecology*, 1147-150.
Daneman E. (1961). “Imipramine in Office Management of Depressive Reactions.” *Diseases of the Nervous System*, 22: 213–217.
David, P. and Dasgupta, P. (1994). "Towards a New Economics of Science." *Research Policy*, 23: 487–521.
Davidson R (1986). "Source of funding and outcome of clinical trials." *Journal of General Internal Medicine*, 1: 155–158.
Davies, N. (2000). Cuckoos, Cowbirds, and Other Cheats. London: T. and A.D. Poyser.
——— and Brooke, M. (1988). "Cuckoos versus reed warblers: Adaptations and counteradaptations". *Animal Behaviour*, 36: 262-284.

——— and Brooke, M. (1989). "An experimental study of co-evolution between the cuckoo Cuckoo canorus and its hosts. I Host egg discrimination." *Journal of Animal Ecology*, 58: 207-224.
Davis, N.S. (1902). "Need of Much More Accurate Knowledge…" *JAMA,* 38*:* 1415-1416.
Dawkins, R. and Krebs, J. (1979). "Arms races between and within species." *Proceedings of the Royal Society of London*, Series B, 205: 489-511.
Day, M. (1998). "Mostly in the mind." *New Scientist*, 159: 13.
Douglas, H. (2009). *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.
Dunn, E. (1985). "Cuckoo: Social pattern and behavior." In: *The Birds of the Western Paleartic, Volume IV.* (ed.) S. Cramp. pp. 407-412. Oxford: Oxford University Press
Dutton, D.(1987). *Worse than the Disease: Pitfalls of Medical Progress*. Cambridge: Cambridge University Press.
Dykstra, D. (1955). "The medical profession and patent and proprietary medicines during the 19th century" *Bulletin of the History of Medicine,* 29: 401-419.
Eisler, P., Morrison, B., and Vanden Brook, T. (2007). "Pentagon Balked at Pleas From Officers in Field for Safer Vehicles: Iraqi Troops Got MRAPs; Americans Waited," USA Today, July 16, A1.
Elkin, I., Shea, T., Watkins, J., Imber, S., Sotsky, S., Collins, J., ... Parloff, M. (1989). "National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments." *Archives of General Psychiatry,* 46: 971-82.
Elliott, C. (2010). *White Coat Black Hat: Adventures on the Dark Side of Medicine*. Boston, Beacon Press.
Elliott, K. (2011). *Is a Little Pollution Good for You?: Incorporation Societal Values in Environmental Research*. New York: Oxford University Press.
Emlen, D. (2014). *Animal Weapons: The Evolution of Battle*. New York: Henry Holt and Co.
——— and Phillips, T. (2006). "Phylogenetic evidence for an association between tunneling behavior and the evolution of horns in dung beetles (Coleoptera: Scarabaeidae: Scarabaeinae)." *Coleopterists Society Monographs*, 5: 47-56.
Enserink, M. (1999). "Can the placebo be the cure?" *Science,* 284: 238–240.
Erwin, S. (2007). "Army Predicts Long Life for Humvee." *National Defense Magazine.* <Online Archive>
Evans, I., Thornton, H., and Chalmers, I. (2006). Testing Treatments: Better Research for Better Healthcare. London: Pinter and Martin.
Fink, D. and Howell I. (2000)." How does Cisplatin kill cells?" In: Platinum-based Drugs in Cancer Therapy. (eds.) I. Kelland and N. Farrell p. 149. Totowa: Humana Press.
Fisher, S. and Greenberg, R. (1989). *The limits of biological treatments for psychological distress: Comparisons with psychotherapy and placebo*. Hillsdale, NJ: Earlbaum.
———, and Greenberg, R. (1997). From Placebo to Panacea: Putting Psychiatric Drugs to the Test. New York: John Wiley and Sons.
Fitzgerald, G. and Patrono, C. (2001). "The coxibs, selective inhibitors of cyclooxygenase-2." The New England Journal of Medicine, 345:433-442.
Flexner, A. (1910). *Medical Education in the United States and Canada: A Report to the Carnegie Foundation for the Advancement of Teaching.* Boston: Merrymount Press.
——— . (1960). *An Autobiography.* New York: Simon and Shuster, pp. 74-75.
Fraga, R. (1998). "Interactions of the parasitic Screaming and Shining Cowbirds (Molothrus rufoaxillaris and M. Bonariensis) with a shared host, the Bay-winged Cowbird (M. badius)." In Parasite Birds and Their Hosts. (ed.) S.I. Rothstein and S.K. Robinson. pp. 173-193. Oxford: Oxford University Press.
Freedman, L. and Schatzkin, A. (1992). "Sample size for studying intermediate endpoints within intervention trials or observational studies." *American Journal of Epidemiology,* 136: 1148-59.

French, J. (1956). "A Formal Theory of Social Power." *Psychological Review,* 63: 181–94.
Fried, S. (1998). *Bitter Pills*. New York: Bantam Books.
Friedhoff, L. (2009). *New Drugs: An Insider's Guide to the FDA's New Drug Approval Process*. Pharmaceutical Special Projects Group. New York, NY
Friedman A. (1975). "Interaction of Drug Therapy with Marital Therapy in Depressive Patients." *Archives of General Psychiatry*, 32: 619–637.
———, Granick, S., Cohen, H., and Cowitz, B. (1966). "Imipramine (Tofranil) vs. Placebo in Hospitalised Psychotic Depressives." *Journal of Psychiatric Research,* 4: 13–36.
Fugh-Berman, A. (2005). "The corporate coauthor." *Journal of General Internal Medicine*, 20: 546-48.
Gabriel, J. M. (2014). Medical Monopolies. Chicago: The University of Chicago Press.
Galula, D. (1964). *Counterinsurgency Warfare: Theory and Practice*. London: Praeger Security International.
Gayl, F. (2008). "Mine Resistant Ambush Protected Vehicle: Ground Combat Element Advocate Science and Technology Advisor Case Study."
Gelenberg, A. (1999). "The editor responds." *Journal of Clinical Psychiatry*, 60: 122.
Gibson, M. (2006). "Attacks on systematic reviews." *US Cochrane Center Conference*, Baltimore, July 14.
Glennmullen, J. (2000). *Prozac Backlash: Overcoming the Dangers of Prozac, Zoloft, Paxil, and Other Antidepressants with Safe Effective Alternatives.* New York: Simon and Schuster.
Gold, H. (1973). "Biographic Essay: Robert Hatcher," In: *Dictionary of American Biography*, American Council of Learned Societies, sup 3: 1944-1945.
Golden, G., Parochka, J., and Overstreet, K. (2002). "Medical education and communication companies: An updated in-depth profile." Journal of Continuing Education in the Health Professions, 22: 55-62.
Goldman, A. (1999). *Knowledge in a Social World*. New York: Oxford University Press.
——— (2011). "Systems Oriented Social Epistemology." In *Social Epistemology: Essential Readings*, (eds.) A. Goldman and D. Whitecomb, pp. 11-37. Oxford: Oxford University Press.
Goodman, L. (1964). "The problem of drug efficacy: An exercise in dissection." In: *Drugs in Our society*. (ed) P. Talalay. pp. 203-213. Baltimore: The Johns Hopkins Press.
Goozner, M. (2004). *The $800 Million Dollar Pill: The Truth Behind the Cost of New Drugs*. Berkeley: University of California Press.
Government Accountability Office. (2012). "Counter-Improvised Explosive Devices: Multiple DOD Organizations are Developing Numerous Initiatives." GAO-12-861R.
Graboys, T., Lown, B., Podrid, P. and DeSilva, R. (1982). "Long-term survival of patients with malignant ventricular arrhythmia treated with antiarrhythmic drugs." The American Journal of Cardiology, 437-443.
Greenberg R. and Fisher, S. (1989). "A second opinion: rethinking the claims of biological Psychiatry," In: The Limits of Biological Treatments for Psychological Distress: Comparisons with Psychotherapy and Placebo. (eds.) S. Fisher, and R. Greenberg. pp 309–336. Hillsdale, NJ: Lawrence Erlbaum Associates.
———, Bornstein, R., Greenberg, M., and Fisher, S. (1992). "A meta-analysis of antidepressant outcome under 'blinder' conditions." Journal of Consulting and Clinical Psychology, 60: 664–69.
———, Bornstein, R., Zboroski, M., Fisher, S., and Greenberg, M. (1994). "A meta- analysis of fluoxetine outcome in the treatment of depression." Journal of Nervous and Mental Disorders, 182: 547-551.
Griffith, D. (1999). "Reasons for not seeing drug representatives." *British Medical Journal*, 319: 69-70.

Grünbaum, A. (1986). "The placebo concept in medicine and psychiatry." *Psychological Medicine,* 16: 19-38.
Harris, G. (2002). "Prilosec's makers switches users to Nexium, thwarting generics." *The Wall Street Journal*, June 6, 2002.
Harrison, M. and Broom, M. (2009). "A game-theoretic model of interspecific brood parasitism with sequential decisions." *Journal of Theoretical Biology*, 256: 504-517.
Hatcher, R. (1916). "The Duty of the Medical Profession to the Council on Pharmacy and Chemistry," JAMA, 67: 1339-1342.
Healy, D. (2004). *Let Them Eat Prozac: The Unhealthy Relationship Between the Pharmaceutical. New York: New York University Press.*
——— (2006). 'Did regulators fail over selective serotonin reuptake inhibitors? *British Medical Journal*, 333: 92-95.
——— (2012). *Pharmageddon*. Berkeley: University of California Press.
——— and Cattell D. (2003). "Interface between authorship, industry and science in the domain of therapeutics." *British Journal of Psychiatry*, 183: 22-27.
Hearings Before Senate Subcommittee on Antitrust and Monopoly (1961). Washington, D.C.: US Government Printing.
Hewlett, A.W. (1913). "Clinical Effects of 'Natural' and 'Synthetic' Sodium Salicylate," JAMA 61*,* 319-320.
Hilts, P. (2003). *Protecting America's Health: The FDA, Business, and One-hundred Years of Regulation.* New York: Alphred A. Knopf.
Hine, L., Laird, N., Hewitt, P., and Chalmers, T. (1989). "Meta-analysis of empirical long-term antiarrhythmic therapy after myocardial infarction." JAMA, 262: 3037-40.
Hollister, L., Overall, J., Johnson, M., Pennington, V. , Katz, G., and Shelton, J. (1964). "Controlled Comparison of Imipramine, Amitriptyline and Placebo in Hospitalised Depressed Patients." *Journal of Nervous and Mental Disease*, 139: 370–375.
Hollon, S., Thase, M., and Markowitz, J. (2002). "Treatment and Prevention of Depression" *Psychological Science in the Public Interest*. Nov, pp.39-77.
Holmbeck, G. (1997). "Toward terminological, conceptual, and statistical clarity in the study of mediators and moderators: Examples from the child-clinical and pediatric psychological literatures." *Journal of Consulting and Clinical Psychology,* 65: 599-610.
Horgan, J. (1999)."Placebo nation." *New York Times*, March 21, 1999, section 4, p 15.
Horton, R. (2004). "The dawn of McScience," *The New York Review of Books,* 51, no 4, pp. 7-9.
Horwitz, A. and Wakefield, J. (2007). *The Loss of Sadness.* Oxford: Oxford University Press.
Howick, J. (2012). *The Philosophy of Evidenced-Based Medicine.* West Sussex: British Medical Journal Books
Hussain, Z. (1970). "Drugs in depressive illness." *British Medical Journal*, 1: 482.
Jackson, M. (2005). "A Survey of Models of Network Formation: Stability and Efficiency." In *Group Formation in Economics: Networks, Clubs and Coalitions,* (eds.) G. Demange and M. Wooders, pp. 11-57. Cambridge: Cambridge University Press.
Jacobi, A. (1906). "Proprietary Medicines," *JAMA, 4*7: 977-978.
Johnson, V. (1947). "The Council on Medical Education and Hospitals," In: *A History of the American Medical Association 1847 to 1947,* (Ed.) M. Fishbein, p. 887-922. Philadelphia: W. B. Saunders.
Josephson, M. (1986). "Treatment of Ventricular Arrhythmias After Myocardial Infarction." *Circulation,* 74:653-658.
Kaptchuk, T. (1998). "Intentional Ignorance: A History of Blind Assessment and Placebo Controls in Medicine," *Bulletin of the History of Medicine,* 72*:* 389-433.

Kassirer, J. (2005). *On the take.* Oxford, UK: Oxford University Press
Keller, M.B., Ryan, N.D., Strober, M., Klein, R.G., Kutcher, S.P., Birmaher, B., ... McCafferty J.P. (2001). "Efficacy of Paroxetine in the Treatment of Adolescent Major Depression: A Randomized Clinical Trial," *Journal of the American Acadamy of Child and Adolescent Psychiatry,* 40: 762-72.
Kirsch, I. (2010). *The Emperor's New Drugs.* New York: Basic Books.
——— and Sapirstein, G. (1998). "Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medication." Prevention and Treatment.
Kitcher, P. (2001). *Science, Truth, and Democracy*. New York: Oxford University Press.
——— (2011). *Science in a Democratic Society.* Amherst: Prometheus Books.
Knox, K., Adams, J., Djulbegovic, B.Knox, Stinson, T.J., Tomor, C., Bennett, C.L. (2000). "Reporting and dissemination of industry versus non- profit sponsored economic analyses of six novel drugs used in oncology." *Annals of Oncology*, 11: 1591-95.
Koepp, R. and Miles, S. (1999). "Meta-analysis of tacrine for Alzheimer disease: The influence of industry sponsors." *JAMA*, 281: 2287-88.
Kondoro, W. and Sibbald, B. (2004) "Drug company Experts Advised Staff to Withhold Data about SSRI Use in Children." *Canadian Medical Association Journal*, 170: 783.
Kourany, J. (2011). "Integrating the Ethical into Scientific Rationality." In: *Science in the Context of Application: Methodological Change, Conceptual Transformation, Cultural Reorientation,* (eds.) M. Carrier and A. Nordmann, *Boston Studies in the Philosophy of Science*, vol. 274, Springer, 2011.
Krepinevich, A. and Wood, D. (2007). "Of IEDs and MRAPs: Force Protection in Complex Irregular Operations. Center for Strategic and Budgetary Assessments."
Krimsky, S. (2003). *Science in the private interest: Has the lure of profits corrupted medical research?* Landham, MD: Rowman and Littlefield
Kuhn, R. (1958). "The treatment of Depressive States with G-22355 (Imipramine hydrochloride)." *American Journal of Psychiatry,* 115: 459-464.
Lamb, C., Schmidt, M., and Fitzsimmons, B. (2009). "MRAPs, Irregular Warfare, and Pentagon Reform." *Institute for National Strategic Studies Occasional Paper 6*. Washington, D.C.: National Defense University Press.
Langston, Nancy. (2010). *Toxic Bodies.* New Haven: Yale University Press
Lapius, S. (1918). "Fellows' Syrup, and Other Preparations of the Hypophosphites," *JAMA,* 70*:* 478-479.
Lasagna, L. (1960). Note to the Paper "The Influence of Side-effects on the Reporting of Symptoms,' by
——— (1962). *The Doctor's Dilemmas.* New York: Harper and Brothers.
——— (1964). "On Evaluating Drug Therapy: The Nature of the Evidence." In *Drugs in Our Society. (e*d.) P. Talay. pp. 91-106. Baltimore: The Johns Hopkins University Press.Letemendia, F. J. J. and Harris, A. D.," *Psychopharmacologia,* 1: 351.
Letemendia, F. and Harris, A. (1959). "The influence of side effects on the reporting of symptoms." *Psychopharmacologia,* 1: 39-47.
Lexchin, J. (1989). "Doctors and detailers: therapeutic education of pharmaceutical promotion?" *International Journal of Health Services,* 19: 663-679.
———, Bero, L., Djulbegovic, B., and Clark, O. (2003). "Pharmaceutical Industry Sponsorship and Research Outcome and Quality." *British Medical Journal,* 326:1167–77.
Leyburn, P. (1967). "A critical look at antidepressant drug trials." *Lancet,* ii: 1135-1137.
Lilienfeld, A. M. (1982). "Ceteris paribus: The evolution of the clinical trial," *Bulletin of the History of Medicine,* 56*:* 1-18.

Lotem, A. (1993). "Learning to recognize nestlings in is maladaptive for cuckoo Cuculus canorus hosts." *Nature* 362: 743-745.
Lown, B. (1979). "Sudden cardiac death: The major challenge confronting contemporary cardiology." *The American Journal of Cardiology*, 43: 313-28.
Lundh, A., Sismondo, S., Lexchin, J., Busuioc, O. and Bero, L. (2012). "Industry sponsorship and research outcome." Cochrane Database of Systematic Reviews, 12.
MacKinnon, D., Lockwood, C., Hoffman, J., West, S., Sheets. V. (2002). "A comparison of methods to test mediation and other intervening variable effects." *Psychological Methods,* 7: 83-104.
Maddy, P. (2007). *Second Philosophy: A Naturalistic Method.* New York: Oxford University Press.
Mandelkern, M. (1999). "Manufacturer support and outcome." *Journal of Clinical Psychology*, 60: 122-23.
Margraf, J., Ehlers, A., Roth, W. T., Clark, D. B., Sheikh, J., Agras, W. S., and Taylor, C. B. (1991). "How 'blind' are double-blind studies?" *Journal of Consulting and Clinical Psychology*, 59: 184–187.
Marks, H. (1997). *The Progress of Experiment: Science and Therapeutic Reform in the United States 1900-1990.* Cambridge: Cambridge University Press.
Matthews, J. R. (1995). *Quantification and the Quest for Medical Certainty.* Princeton: Princeton University Press.
Mayo, D. and Spanos, A. (2010). Error and Inference: Recent Exchanges on Experimental Reasoning, Reliability, and the Objectivity and Rationality of Science. New York: Cambridge University Press.
McTavish, J. R. (2004). *Pain and Profits: the History of the Headache and Its Remedies in America.* New Brunswick: Rutgers University Press.
Medical letters. (1986). "Tambocor." *Medical Letters*, 706: 20.
——— (1987). "Encainide." *Medical Letters*, 734: 52.
Meigs, M. (1990). *Slide Rules and Submarines: American Scientists and Subsurface Warfare in World War II*. Washington D.C.: National Defense University Press.
——— (2010). "IEDs: Weapon of Choice for Insurgents from Iraq to Afghanistan." Speech given at New America. Apr 12, 2010.
Merton, R. (1942). Science and Technology in a Democratic Order," *Journal of Legal and Political Sociology* 1 (1942), 115-126.
Meyers, Robert. (1983.) *DES: The Bitter Pill.* New York: Seaview/Putnam.
Mintz, M. (1965). *The Therapeutic Nightmare: A Report on the United States Food and Drug Administration, The American Medical Association, Pharmaceutical Manufacturers, and Others in Connection with the Irrational and Massive Use of Prescription Drugs that may be Worthless, Injurious, or Even Lethal.* Boston: Houghton Mifflin.
Mirowski, P., and Van Horn, R. (2005). "The Contract Research Organization and the Commercialization of Scientific Research." *Social Studies of Science,* 35: 503–548.
Moksnes, A., Røskaft, E., and Korsnes, L. (1995). "On the evolution of blue cuckoo eggs in Europe." *Journal of Avian Biology*, 26: 13-19.
Moncrieff, J., Wessely, S., and Hardy, R. (1998). "Meta-analysis of trials comparing antidepressants with active placebos." *British Journal of Psychiatry,* 172: 227–231.
Montejo-González, G., Llorca, J., Izquierdo, J., Ledesma, A., Bousono, M., Calcedo, A., ... Vicens, E. (1997). "SSRI-induced Sexual Dysfunction: Fluoxetine [Prozac], Paroxetine [Paxil], Sertraline [Zoloft], Fluvoxamine [Luvox], in Prospective Multicenter and Descriptive Clinical Study of 344 Study." *Journal of Sex and Marital Therapy,* 23: 176-94.

Montgomery, J., Byerly, M., Carmody, T., Li, B., Miller, D.R., Varghese, F., and Holland, R. (2004). "An analysis of the effect of funding source in randomized clinical trials of second generation antipsychotics for the treatment of schizophrenia." *Controlled Clinical Trials,* 25: 598-612.

Moore, T. (1995). *Deadly Medicines: Why Tens of Thousands of Heart Patients Died in America's Worst Drug Disaster.* New York: Simon and Schuster

Morganroth, J. (1984). "Premature Ventricular Complex." *JAMA,* 252: 673-676.

——— and Horowitz, L. (1984). "Flecainide: Its proarrhythmic effect and expected changes on the surface electrocardiogram." *American Journal of Cardiology,* 53: 89B-94B

Morris, J., and Beck, A. (1974). "The efficacy of antidepressant drugs. A review of research (1958-1972)." Archive of General Psychiatry, 30: 667-74.

Moseley, J. B., O'Malley, K., Petersen, N., Menke, T., Brody, B., Kuykendall, D., … and Wray,N. (2002). "A Controlled Trial of Arthroscopic Surgery for Osteoarthritis of the Knee." *New England Journal of Medicine,* 347:81-88.

Moynihan, R. (2003). "Who pays for the pizza? Redefining the relationship between doctors and drug company. Part I, Entanglement and Part 2, Disentanglement." *British Medical Journal,* 326: 1189-1196.

——— (2010). "Drug firm's legal action against government advisers is described as 'intimidation'." *British Medical Journal,* 341: c6527.

——— and Cassels, A. (2006). Selling Sickness: How the World's Biggest Pharmaceutical Companies are Turning Us All into Patients. New York: Nation Books.

Mukharji J., Rude R., Poole K., Croft, C., Thomas, L.J., Jr., Strauss, H.W., ... Cooperating investigators Multicenter Investigation of the Limitation of Infarct Size. (1982). "Late sudden death following acute myocardial infarction, importance of combined presence of repetitive ventricular ectopy and left ventricular dysfunction." *Clinical Research,* 30: 108A.

Nagl, J. (2003). *Learning to Eat Soup with a Knife: Counterinsurgency Lessons From Malaya and Vietnam.* Chicago: University of Chicago Press

Nash, M. (1962). "The double blind procedure: Rationale and empirical evaluation." *Journal of Nervous and Mental Disease,* 134: 34-7.

Oldani, M. (2004). "Thick prescriptions: Toward an interpretation of pharmaceutical sales practices." *Medical Anthropology Quarterly,* 18: 325-356.

Oreskes, N. and Conway, E. (2010). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming.* New York: Bloomsbury Press.

Orlowski, J. and Wateska, L. (1992). "The effects of pharmaceutical firm enticements on physician prescribing patterns: there's no such thing as a free lunch." *Chest,* 102: 270-273.

Orwell, G. (1945). "You and the Atomic Bomb." *Tribune.* October 19.

——— (1949). *Nineteen Eighty-Four.* New York: Harcourt Brace Jovanovich.

Osborn, K. (2008). "MRAP Breakthrough."*Defense News,* October 6, 11.

Oxtoby A., Jones, A., and Robinson, M. (1989). "Is your 'double-blind' study truly double- blind?" *British journal of Psychiatry,* 155: 700-701.

Parascandola, J. (1992). *The Development of American Pharmacology: John J. Able and the Shaping of a Discipline*. Baltimore: The Johns Hopkins Press.

Payne, R. (1974)." The evolution of clutch size and reproductive rates in parasitic cuckoos," *Evolution,* 28: 169-181.

Popper, K. (1963). *Conjectures and refutations.* London: Routledge and Keagan Paul Ltd.

Puckner, W. (1919). "The Council on Pharmacy and Chemistry, Present and Future," *JAMA,* 72*: 1339-1343.*

——— and Clark, A. (1908). “Burnham’s Soluble Iodin,” JAMA, *50:* 1055.
——— and Hilpert, W. (1908). “Zyme-oid,” JAMA, 50: 1706.
——— and Hilpert, W. (1908). “Uriseptin: Another Proprietary of the 'Hidden Drug' Type,” JAMA, ——
——— and Warren, L. (1910). “Noitol and Anadol,” JAMA, 54: 1704-170551: 773-774.
Quintiles. (2014). Delivering Real-world Evidence to Demonstrate Product Safety and Value [Bochure]. N.P:n.p.
Quitkin, F. (1999). "Placebos, drug effects, and study design: a clinician's guide." *American Journal of Psychiatry*, 156: 829–836.
———, Rabkin, J., Gerald, J., Davis, J., and Klein, D. (2000). "Validity of Clinical Trials of Antidepressants." *The American Journal of Psychiatry,* 157:327.
Rennie, D. and Flanagin, A. (1994). "Authorship! authorship! guests, ghosts, grafters, and the two-sided coin." *JAMA,* 271: 469–71.
Ridker, P., and Torres, J., (2006). "Reported outcomes in major cardiovascular clinical trials funded by for-profit and not-for-profit organizations: 2000-2005." *JAMA,* 295: 2270-2274.
Robinson, W. (1904). “The Composition of Some So-called Ethical Synthetics and ‘Ethical’ Nostrums,” *JAMA,* 41*:* 1016-1017.
Rochon, P., Gurwitz, J., Simms, R., Fortin, P., Felson, D., Minaker, K., and Chalmers, T. (1994). "A study of manufacturer-supported trials of nonsteroidal anti-inflamatory drugs in the treatment of arthritis." *Archives of Internal Medicine,* 154: 157-163.
Rohlfs, C. and Sullivan, R. (2011). "The Cost-Effectiveness of Armored Tactical Wheeled Vehicles for Overseas U.S. Army Operations." <Online>
Rosen, G. (1983). *The Structure of American Medical Practice, 1875-1941.* Philadelphia: University of Pennsylvania Press.
Rost, P. (2006). *The Whistleblower: Confessions of a Healthcare Hitman.* Brooklyn: Soft Skull Press.
Rumsfeld, D. (2004). A partial transcript can be found at: http://www.pbs.org/newshour/bb/military-july-dec04-questioned_12-8/
Rupniak, N. (2003). "Animal models of depression: Challenges from a drug development perspective." *Behavioural Pharmacology,* 14: 385-390.
Russo, F. and Williams, J. (2004). "Interpreting causality in the health sciences." International Studies in the Philosophy of Science, 21: 1157-1170.
Safer, D. (2002). "Design and Reporting Modifications in Industry-Sponsored Comparative Psychopharmacology Trials." *Journal of Nervous and Mental Disorder,* 190: 583–592.
Salisbury, J. (1906). “The subordination of medical journals to proprietary interests,” *JAMA*, 46, 1337-1338.
Sarkesian, S. (1993). *Unconventional Conflicts in a New Security Era: Lessons From Malaya and Vietnam.* London: Greenwood.
Schulman, K., Seils, D., Timbie, J., Sugarman, J., Dame, L., Weinfurt, K.P. … Califf, R.M. (2002). "A national survey of provisions in clinical trial agreements between medical schools and industry sponsors." *NEJM,* 347: 1335-41.
Schulz, K., Altman, D., Moher, D.,and Fergusson, D. (2010). *CONSORT 2010 "C*hanges and testing blindness in RCTs." *Lancet,* 375: 1144-46.
Schwartz, S. (1998). *Atomic Audit: The Costs and Consequences of U.S. Nuclear Weapons since 1940.* Washington D.C.: Brookings Institution Press.
Segraves, R., Segraves, K., and Bubna, C. (1995). “Sexual Function in Patients Taking Bupropion [Wellbutrin] Sustained Release,” *Journal of Clinical Psychiatry,* 56: 374.

Servedio, M. and Hauber, M. (2006). "To eject or to abandon? Life history traits of hosts and parasites interact to influence the fitness payoffs of alternative anti-parasite strategies." *Journal of Evolutionary Biology,* 19: 1585-1594.
Shooter, M. (2004). "Taken on Trust," *BBC Original* airdate: 3:10:04
Simmons, G. (1906). " Proprietary Medicine: Some General Considerations," *JAMA,* 46*:* 1333-1337.
——— (1907). "The Commercial Domination of Therapeutics and the Movement for Reform," *JAMA,* 48*:* 1645.
Sinclair, U. (1935/1994). "*I, Candidate for Governor: And How I Got Licked.* Berkeley: University of California Press, 1935/1994), 109
Sinclair, W. (1996). "Answering the Landmine." *Marine Corps Gazette,* 80, no. 7, pp. 37-40.
Sismondo, S. (2009). "Ghosts in the Machine: Publication Planning in the Medical Sciences." *Social Studies of Science,* 39: 171-198.
Smith, R. (2005). "Medical journals are an extension of the marketing arm of pharmaceutical companies." *PLoS Medicine,* 2: 364-66.
Sollmann, T. (1908). The Broader Aims of the Council of Pharmacy of the American Medical Association. Chicago: AMA.
——— (1913). "Yesterday, To-day and To-morrow: The Activities of the Council on Pharmacy," *JAMA,* 61*:* 6.
——— (1917). "The Crucial Test of Therapeutic Evidence," *JAMA,* 69*:* 199.
Spiegel, A. (2010). "How to win doctors and influence prescriptions." *NPR* Oct. 21.
Stelfox, H., Chua, G., O'Rourke, K., and Detsky. (1998). "Conflicts of interest in the debate over calcium-channel antagonists." *NEJM,* 338: 101-6.
Stewart, F. (1901). "Proposed Nation Bureau of Materia Medica," *JAMA,* 36: 1177*.*
Straus, S., Richardson, W., Glasziou, P., and Haynes R. (2005) Evidence-Based Medicine: How to Practice and Teach EBM (Third Edition), Edinburgh: Churchill Livingstone.
Swann, J. (1988). *Academic Scientists and the Pharmaceutical Industry: Cooperative Research in the Twenty-first Century.* Baltimore: The Johns Hopkins University Press.
Takasu, F. (1998). "Why do all host species not show defense against avian brood parasitism: Evolutionary lag or equilibrium?" *American Naturalist*, 151: 193-205.
Talbot, M. (2000). "The placebo prescription." *The New York Times.* January 9, Section 6; p 34.
The Medical Brief. (1910). "The Blind Infatuation of the A.M.A. Leaders," The Medical Brief, 38:565-567
The Medical Standard. (1911). "A Critical Laboratory" The Medical Standard, 34: 42.
Thomson, R. (1982). "Side effects and placebo amplification." *The British Journal of Psychiatry*, 140: 64-68.
Uhlenhuth E, and Park, L. (1964). "The Influence of Medication (Imipramine) and Doctor in Relieving Depressed Psychoneurotic Outpatients." *Journal of Psychiatric Research*, 2: 101–122.
United States of America ex. rel. Christopher R. Gobble v. Forest Pharmaceuticals, No. 03-10395-NMG. United States District Court, District of Massachusetts. 2010.
USMC. (2005). *Vehicle Hardening Information Brief* <Online>.
Vandenbroucke, J., Helmerhost, F., and Rosendaal, F. (2000). "Competing interests and controversy about third generation oral contraceptives." *BMJ*, 320: 381-2.
Velebit, V., Podrid, P., Lown, B., Cohen, B., and Graboys T. (1982). "Aggravation and provocation of ventricular arrhythmias by antiarrhythmic drugs." *Circulation*, 65: 886-894.
von Clausewitz, C. (1832/1989). *On War*. Princeton: Princeton University Press.

von Knorring, A., Olsson, G., Thomsen, P., Lemming, O., and Hultén, A. (2006). "A randomized, double blind, placebo-controlled study of citalopram in adolescents with major depressive disorder." *Journal of Clinical Pharmacology*, 26: 311-315.

Wagner, K., Robb, A., Findling, R., Lin, J., Gutierrez, M., and Heydorn, W. (1994). "A Randomized, Placebo-Controlled Trial of Citalopram for the Treatment of Major Depression in Children and Adolescents." *American Journal of Psychiatry,* 161: 1079-1083.

Walker, P., Cole, J., Gardner, A., Hugh, A., Johnson, J., Batey, S.R., and Lineberry, C.G. (1993). "Improvement in Fluoxetine [Prozac]-Associated Sexual Dysfunction in Patients Switched to Bupropion [Wellbutrin]," *Journal of Clinical Psychiatry,* 54: 459-465.

Wazana, A. (2000). "Physicians and the pharmaceutical industry: is a gift always a gift?" *JAMA,* 283: 373-380.

Weintraub W, and H. Aronson. (1963). "Clinical Judgment in Psychopharmacological Research." *Journal of Neuropsychiatry* 5: 65–70.

Whitaker, R. (1998). "Lure of Riches Fuels Testing." *Boston Globe,* November, 17.

——— (2002). *Mad in America. Bad Science, Bad Medicine, and the Enduring Mistreatment of the Mentally Ill*. Cambridge: Perseus.

White, K., Kando, J., Park, T., Waternaux, C., and Brown, W. (1992). "Side effects and the 'blindability' of clinical drug trials." *Clinical Research Reports,* 149: 1730–31.

Wilson I, J. Vernon, T. Guin, and M. Sandifer. (1963). "A Controlled Study of Treatment of Depression." *Journal of Neuropsychiatry*, 4: 331-337.

Wilson, C. (2006). "Improvised Explosive Devices (IEDs) in Iraq: Effects and Countermeasures." *Congressional Research Service Report to Congress,* RS22330. Washington, D.C.

Winkle, R., Mason, J., Griffin, J., and Ross, D. (1981). "Malignant ventricular tachyarrhythmias associated with the use of encainide." *American Heart Journal,* 102:857-864.

Woosley, R. (1990). "CAST: Implications for drug development." *Clinical Pharmacology and Therapeutics,* 47: 553-556.

Worral, J. (2002). "What Evidence in Evidence-Based Medicine?" *Proceedings of the Philosophy of Science Association, p.* S316-S330.

——— (2007). "Evidence in medicine and evidence-based medicine," *Philosophy Compass*, 2: 981-1022

——— (2007a). "Evidence in medicine and evidence-based medicine," *Philosophy Compass,* 2: 981-1022.

——— (2007b). "Why there's no cause to randomize." *British Journal for the Philosophy of Science,* 58:451-488.

——— (2010). "Evidence: Philosophy of science meets medicine." *Journal of Evaluation in Clinical Practice*, 16: 356-362.

Young, A. (2001). "Recurrent unipolar depression requires prolonged treatment." *British Journal of Psychiatry,* 178: 294-295.

Young, J. H. (1961). *The Toadstool Millionaires: A Social History of Patent Medicines in America Before Federal Regulation.* Princeton: Princeton University Press.

Young, S. (1996). "The use of placebos in psychiatry: a response to the draft document prepared by the TRI-Council Working Group." *Journal of Psychiatry and Neuroscience,* 21: 235–38.

Zollman, Kevin. (2007). "The Communication Structure of Epistemic Communities." *Philosophy of Science,* 74: 574–87.

——— (2010). "The Epistemic Benefit of Transient Diversity." *Erkenntnis,* 72: 17–35.

——— (2013). "Network Epistemology: Communication in Epistemic Communities." *Philosophy Compass,* 8: 15–27.

Zorpette, G. (2008) "Countering IEDs," *IEEE Spectrum* <Online>

APPENDIX A: Statistical Analyses of an Antidepressant Trial

This data set was obtained by contacting Dr. Elkin, the principal investigator of the Treatment of Depression Collaborative Research Project, a multi-site study funded by the National Institute of Mental Health. Results here compare the groups receiving imipramine (IMI-CM) to the placebo (PLA-CM) and use the Hopkins Symptom Checklist-90 (HSCL-90) as the measure of improvement. The first two analyses use the end point 204 sample to make use of all the available data in establishing general phenomena (PLA-CM, n = 35; IMI-CM, n =46). The third and fourth analysis will be restricted to the 71 patients who completed the study (PLA-CM, n = 35; IMI-CM, n =36) to assess previous findings. One case was deleted from the fourth analysis on the basis of the regression diagnostics (it was overly influential on the regression). Further details on the sample can be found here (Elkin, *et al.,* 1989).

Statistical Analyses

The first analysis uses a binomial distribution to assess whether participants and/or raters can identify group assignment above chance levels. Results are similar in the eighth and sixteenth weeks. The former are reported. The second analysis uses logistic regression to determine whether side effects and/or improvement influence beliefs about group assignment. Results are similar in both the eight and sixteenth weeks. The latter are reported. The final analysis employs a series of ordinary least squares (OLS) regressions to determine whether belief about group assignment mediates the effect of imipramine. The t-statistic is calculated by the method described in Freedman and Schatzkin (1992) as per recommendations showing it to be the most reliable test of mediating effects (MacKinnon *et al*. 2002).

Results

Failure of blinding procedures

Of the patients receiving imipramine, 26/27 believed they were receiving the drug. Patients receiving placebo guessed roughly at chance (10/19). Combined, patients had correct beliefs 78% of the time ($p < .001$).

Side effects predict patient beliefs, improvement does not

Side effects ($p < .001$), but not improvement ($p = .14$, n.s.) influences which group patients believe they are in. Side effects continue to significantly predict patients' beliefs after improvement has been removed from the model ($p < .001$). The model's fit cannot be rejected using a Hosmer & Lemeshow test for lack of fit ($\chi^2 = 7.83$, df=7 $p = .351$). In cases where improvement was an independent predictor of patient belief, SEM might be used in place of the analysis below.

The effect of imipramine is completely mediated by patient belief

Let A be the treatment condition, B be the group the patient believes themselves to be in, and C be symptomatic improvement. First, A → C is assessed to determine whether treatment group affects improvement. An OLS regression shows that there is a significant difference between treatment conditions (see table 1). Next, the paths between A → B and B→ C are assessed independently. Again, each path is statistically significant. Finally, both A and B are put into the model. The t-statistic for mediation is calculated with the equation below (Freedman and Schatzkin, 1992):

$t_{N-2} = \frac{\tau - \tau'}{\sqrt{\sigma_\tau^2 + \sigma_{\tau'}^2 - 2\sigma_\tau \sigma_{\tau'} \sqrt{1-\rho^2}}}$, where τ is the β from the regression with just the treatment group, τ' is β from the regression with both variables in the model, and ρ is the correlation between treatment group and believed group ($r = .553$). Entering in the observed values into the equation, we get

t (34) = 2.11, p = .042. Thus, the null hypothesis that τ - τ'= 0 (viz. the effect of treatment group is equivalent when side effects are added into the model) is rejected.

MacKinnon *et al*. (2002) differentiate between complete and partial mediation. Complete mediation occurs when β' is not significantly different from 0, a partial mediation occurs when the test for mediation is significant, but β' remains significant. The hypothesis that β' = 0 is evaluated in the last model, and the hypothesis cannot be rejected (p = .486). Thus, this presents a case of complete mediation. That is, the benefit of the antidepressant disappears once we take into account which group patients believed themselves to be in.

Table 1—Mediator Analysis of Patient Predictions

Predictor	β	*SE* β	t	*df*	*p*	R^2
			A → C			
Treatment Group	0.318	0.146	2.181	35	0.036	0.123
			A → B			
Treatment Group	0.486	0.126	3.87	35	<.001	0.306
			B → C			
Believed Group	0.48	0.157	3.058	35	0.004	0.216
			(A & B) → C			
Treatment Group	0.122	0.167	0.784	34	0.486	(n.s.)
Believed Group	0.406	0.19	2.125	34	0.041	
Model						0.228