David M. Allen • James W. Howell
Editors

# Groupthink in Science

## Greed, Pathological Altruism, Ideology, Competition, and Culture

Springer

*Editors*
David M. Allen
Department of Psychiatry
University of Tennessee Health
Science Center
Memphis, TN, USA

James W. Howell
Department of Psychiatry
University of Tennessee Health
Science Center
Memphis, TN, USA

ISBN 978-3-030-36821-0 ISBN 978-3-030-36822-7 (eBook)
https://doi.org/10.1007/978-3-030-36822-7



This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Chapter 10
# Law Versus Science


**Brian Martin**

## Introduction

Pieter Cohen, a professor of medicine at Harvard University, has carried out research into the effects of nutritional supplements, the sorts commonly consumed by bodybuilders. Cohen published his findings in scientific journals and also publicized them more widely. Following this, Jared Wheat, the owner of the supplement producer Hi-Tech Pharmaceuticals, sued Cohen for defamation (Robbins, 2017).

This might have seemed to be an obvious SLAPP – Strategic Lawsuit Against Public Participation – namely a legal action serving to restrain legitimate participation in matters of public interest. SLAPPs use various torts, most commonly defamation, to scare targets. In US courts, plaintiffs hardly ever win because of the First Amendment right to petition the government, for example, to write letters of complaint to politicians and public officials. However, the point of SLAPPs is seldom to win in court but rather to discourage participation in public matters (Pring & Canan, 1996; Sheldrick, 2014).

Because of the damaging effect of SLAPPs, many US states have passed anti-SLAPP laws. Massachusetts had such a law, so it seemed that Wheat's suit should have been dismissed. However, a judge overruled the state's anti-SLAPP law, saying it prevented the plaintiff from obtaining a trial by jury.

Cohen, backed by Harvard, had to defend in court. They won, but only in legal terms. Defending the case required a considerable amount of time, effort, and emotional energy. Wheat stated that he hoped his legal action would discourage other scientists from making "baseless allegations," and advised scientists to "think twice and do better research, knowing you can get sued if you do this" (quoted in Robbins, 2017). The legal action stimulated commentary by Cohen and others that referred to similar cases and reflected on the inappropriateness of using legal forums for dealing with scientific disagreements (Bagley, Carroll, & Cohen, 2017; Carroll, 2017; Katz & Redberg, 2017).

B. Martin (✉)
Humanities and Social Inquiry, University of Wollongong, Wollongong, NSW, Australia
e-mail: bmartin@uow.edu.au



D. M. Allen, J. W. Howell (eds.), *Groupthink in Science*,
https://doi.org/10.1007/978-3-030-36822-7_10

Wheat's case against Cohen is an example of how legal actions can influence the path of scientific research. Most commonly, they discourage research on specific topics or from particular perspectives.

In this chapter, I describe several ways that laws and legal actions can restrain scientific research. The next two sections describe relevant aspects of science and law. Then, I address the use of defamation actions against researchers, showing the role of domain shifting and illustrating the chilling effect of legal action. Next is a section on laws against euthanasia and how they indirectly discourage research in the area. After this is a discussion of intellectual property and how a legal regime set up to foster innovation can actually constrain it. Finally, I present a framework for analyzing responses to legal constraints on scientific research.

Legal interventions into scientific issues can be interpreted as attempts by self-interested groups to stymie or slant knowledge claims to their own advantage. In some cases, the interventions are from outsiders to the scientific enterprise, while in others, they are from insiders. Most scientists take the easy way out by avoiding research that might or has come under legal attack: for them, maintaining a strife-free career is more important than pursing principled action. Only occasionally do scientists unite against legal threats and constraints.

## *The Domains of Science and Law*

To understand the effect of legal actions on research, it is useful to consider key features of science and law, especially the contrasts between them. Both science and law are enterprises that pursue truth in some sense. Science operates by processes of observation, experiment, theory development, and theory testing, aiming to develop and verify ways of explaining the world, both the natural and the social world. Scientists undertake research and publish it openly, so it is available for scrutiny, subject to critique and potentially the launching pad for further investigations. Decisions about what counts as a scientific fact or a valid theory are made through a sort of consensus process. There is no authority that pronounces the truth or falsity of a claim. Scientific knowledge emerges through collective processes of making claims and counterclaims until all or most researchers agree.

Law is a system of rules, based on statutes or precedents, that are applied to particular cases. Interpretations of the law are made by judges. Law can change, through legislation or reinterpretation. Judgments are published for all to read, but less to question than to understand and show relevance to future cases.

For some purposes, the formal differences between science and law are less important than the fact that they are different domains, each with its own set of procedures, practitioners, criteria, and aims. The aim of science is truth, whereas the aim of law is justice.

A legal action against a scientist, such as Wheat's suit against Cohen, involves movement from one domain to another, namely from science to law. It thus can serve to hinder the normal operation of science, with outcomes such as hindering

particular researchers or discouraging certain types of research. Rather than research being assessed by peers according to scientific criteria, it is assessed by judges or juries according to entirely different sets of rules.

Another important factor in domain shifting is cost. Some research can be expensive in terms of salaries and equipment, but this cost is usually covered by funders – typically universities, governments, or corporations, not by individual scientists. Defending a legal action can be very expensive, involving tens or hundreds of thousands of dollars, and onerous for an individual and sometimes for an organization. There is also a significant cost in terms of time. Although lawyers run a case on behalf of a defendant, often the defendant spends many hours preparing documents. This represents an opportunity cost in research time foregone. Wheat's initial claim against Cohen was for $20 million. Although this might be considered an ambit claim or as a form of intimidation, even a judgment awarding $1 million in damages would be devastating for most individual defendants.

## Constraints on Science

It needs to be said that science is never unfettered. Although truth is the guiding light for researchers, some truths are deemed undesirable. To take an extreme example, studying the effects of nuclear explosions on people and the environment could be done by dropping bombs on populations. Nazi doctors' experimentation on prisoners is considered a crime. Various treaties and laws, for example, on land mines and animal experimentation, constrain research. Some governments have attempted to control research on encryption and stem cells. Studies of vulnerable groups, such as children and prisoners, are limited by the requirements of institutional review boards and ethics committees.

Science is thus constrained in many ways, in part by legal restrictions. The search for truth needs to be undertaken in the context of other values. How then is it possible to assess whether a legal action is a legitimate expression of some public interest or a harmful restraint on the search for knowledge? Various factors can be considered, including widely endorsed principles, involvement by affected parties in deciding on rules affecting science, and examination of who benefits from legal actions. To take two contrasting examples, research to improve methods of torture is in conflict with human rights principles, whereas research on nutritional supplements has the potential to benefit consumers. Hence, legal restraints on torture research can be justified far more easily than legal restraints on supplements research.

In some cases, specific legal restrictions on research are clearly imposed in the service of vested interests. For example, in the United States, the influence of the National Rifle Association is sufficiently great that Congress in 1996 passed a law preventing federal funds for injury prevention at the Centers for Disease Control and Prevention from being used to promote gun control. The result is that gun violence is grossly understudied (Stark & Shah, 2017). However, when laws are used

to inhibit research, often there are other rationales than straight-out censorship. All sorts of laws affect scientific research directly or indirectly, for example, laws on environmental protection, animal welfare, building codes, vehicle safety, minimum wage, employment contracts, pension funds, broadcasting, and monopolies. To illustrate the issues involved, three cases will be examined in more detail: defamation, euthanasia, and intellectual property.

## *Defamation*

Defamation law is commonly seen as an attempt to balance two competing values: protection of reputation and protection of free speech. One person's speech can hurt another person's reputation. When someone feels their reputation has been harmed by another's speech, they can go to court seeking damages, and this very possibility serves to inhibit reputation-damaging speech. On the other hand, defamation suits can inhibit speech that serves the public interest, so the law provides defenses. For example, a defendant may be able to defeat a charge of defamation by demonstrating the truth of statements made. Other defenses include qualified privilege, for example, when a teacher gives grades that hurt a student's reputation, and parliamentary privilege, when an elected representative makes statements in parliament.

Publication of scientific papers, and reports of research, potentially can harm the reputation of individuals, including other scientists whose ideas or contributions are challenged. This sets the stage for invoking defamation law in ways that block or discourage research.

Alex de Blas was a student at the University of Tasmania. For her fourth year of undergraduate study, called the honors year, she wrote a thesis about pollution from the Mt Lyell mine in the state of Tasmania. The owners of the mine threatened de Blas and the university with an action for defamation, demanding that it not be published (Montgomery, 1994).

Hilary Koprowski was a pioneer in developing a vaccine for polio. In a mass test of his vaccine in the late 1950s, it was given to nearly a million people in what today is the Congo. Decades later, a few individuals proposed that this vaccination campaign may have inadvertently led to the emergence of the disease AIDS. Tom Curtis, a journalist for the *Houston Post,* learned about this theory, investigated further, and wrote a story in *Rolling Stone,* generating enormous interest (Curtis, 1992). Koprowski sued Curtis and *Rolling Stone* for defamation. The case was highly expensive, and eventually *Rolling Stone* settled, paying Koprowski $1 and issuing a "clarification." In the discovery phase of the proceedings, Curtis had to provide all his interview notes. He had planned a follow-up article, but this was cancelled. Furthermore, if he had wanted to do further interviews, he would have had to tell interviewees that anything they told him might be accessed in a future legal action. Koprowski's legal action thus had a severe chilling effect on further investigation of the polio-vaccine theory for the origin of AIDS (Martin, 2010).

## *Euthanasia Research*

Research on euthanasia is severely restricted by the law, though not by direct legal action. In Nazi Germany in 1939, Hitler initiated a program of killing people with disabilities, euphemistically called euthanasia, thereby stigmatizing the term for decades to come. After World War II, interest in peaceful death for humanitarian reasons developed largely as a result of medical advances. Technologies such as defibrillators and feeding tubes mean that people who once would have died can be maintained alive, but often with greatly reduced quality of life. Some, suffering greatly from lack of autonomy, indignity, breathlessness, or intractable pain, sought an early death.

In most countries, the means for a violent death abound, including guns, high places, trains, and rope. On the other hand, means for peaceful death have been increasingly limited, with drugs that might provide fatal overdoses being restricted. This has led to a push for legalization of voluntary euthanasia. There are three main options for a peaceful death. The first, called active euthanasia, involves a doctor giving a patient a legal injection. The second, called physician-assisted suicide or physician-assisted dying, involves a doctor giving a patient a prescription for lethal drugs; the patient, if wishing to die, then takes the drugs. The third option, called self-deliverance or do-it-yourself euthanasia, involves a person obtaining lethal drugs or constructing an exit bag and then ending their life, without assistance.

In most countries, it is legal to commit suicide but illegal to help someone to end their life. Most of the writing in the area is about the ethics and legalities of euthanasia with little attention to research. It can be argued that the legal restraints on euthanasia have created a related restraint on research.

Even in places where euthanasia is illegal, it still occurs. For example, sympathetic doctors may covertly give patients access to lethal drugs or give patients lethal injections. To research this practice requires great care. Roger Magnusson (2002) carried out interviews with Australian doctors, documenting an underground euthanasia practice, and revealed that doctors sometimes botched their attempts to help patients die, usually because of lack of knowledge and training. This is one of the few interview-based studies of euthanasia practiced in places where it is illegal. Such research is restrained for two reasons: subjects of the research – doctors who assisted patients to die – are wary of revealing actions that could lead to deregistration or criminal prosecutions, and the research itself is generally unwelcome by governments and medical authorities that oppose legalization of euthanasia.

Russel Ogden, an academic at Kwantlen Polytechnic University in British Columbia, pursued research into assisted dying and do-it-yourself euthanasia. In the course of his studies, he observed several individuals ending their lives (e.g., Ogden, 2010). This sort of investigation, vital to learning how dying intended to be peaceful actually operates and can sometimes go wrong, was not welcomed in some quarters within his university. Ogden encountered obstacles to his research at several universities (Hager, 2015).

Advocates of everyone having the option of a peaceful death have looked for a "peaceful pill," a metaphor for any means by which individuals can reliably end their lives peacefully. The current preferred option is the drug pentobarbital, commonly called Nembutal. It is available at veterinary supply shops in some countries but highly restricted in many. For the "Nutech" group searching for a peaceful pill, one path would be to develop the capacity to synthesize Nembutal cheaply and easily (Côté, 2012). However, efforts along these lines have so far not succeeded, being limited by human and financial resources. Governments do not sponsor this sort of research, except perhaps for classified military and national security purposes.

Laws against euthanasia thus restrain scientific research in several ways. Investigations into the ways doctors and others (such as relatives) covertly end the lives of individuals to end their extreme suffering are rare, in large part due to criminal sanctions against helping others to die. Obstacles include obtaining project approval from research institutions and obtaining the trust of doctors to learn about their practices. A similar difficulty faces those, like Russel Ogden, who study do-it-yourself euthanasia. Finally, research into means for peaceful dying is restrained by stigma and resources.

Some would say that research into euthanasia is undesirable, especially if it helps justify euthanasia, discourages improvement in palliative care, and increases the number of people whose lives are ended prematurely. On the other hand, others believe research into euthanasia can contribute to improved quality of life, for example, by determining how deaths can be more peaceful, by better judging when euthanasia is warranted, and by investigating the potential for abuse.

## Intellectual Property

The term "intellectual property" refers to a variety of laws, including copyright, patents, trademarks, and plant variety rights. Their common feature is putting legal restrictions on the use of ideas. For example, under current copyright law, as soon as a person writes some original words (such as the words in this chapter), they are copyrighted, and others cannot legally reproduce them for profit. Copyright can be assigned to others or sold. Copyright currently lasts for 70 years after the author's death.

The rationale for intellectual property is to stimulate the production of new ideas and devices. It operates by giving a temporary monopoly over the ideas and devices, to enable the creator to gain a benefit. The curious feature of intellectual property is that it legally restrains innovation in the name of stimulating innovation. Unlike material objects, ideas can be used by many people at the same time. Only one person can be wearing a pair of shoes at a time, whereas a poet can enjoy her poem while millions of others are reading it too.

To optimally stimulate the production of new ideas, the duration of protection needs to be adjusted to provide a balance between encouraging the creator (through restraining use by others) and enabling others to build on a creator's work. For

example, novels usually have most of their sales within the first year of publication. Therefore, it can be argued that there is no need for copyright to extend beyond 1 year or perhaps a few years. Very few authors write more or better works by knowing that their copyright extends decades beyond their death. Excessive copyright terms, which keep being extended, restrain creativity by others. A classic example is Mickey Mouse, under copyright to the Disney Corporation. This copyright now does nothing to stimulate greater creativity by the original creator of Mickey Mouse, meanwhile preventing others from using Mickey Mouse for their own creations. Copyright with such a great duration thus serves to restrain innovation. It is a "monopoly privilege" enforced by law (Drahos, 1996).

Scientific research is partially protected from the restraints involved with intellectual property. Scientific papers are copyrighted in the usual way. Initially, the author holds copyright, but many journal publishers ask authors to assign them copyright for the purposes of sales, databases, and other uses. Other authors can quote from published papers as long as the quotes not too long, according to fair use provisions. This allows other authors to quote portions of a text for the purposes of exposition or criticism.

Courts have interpreted copyright of scientific papers as applying only to the expression, namely the words used, not to the ideas. Furthermore, scientific formulas cannot be copyrighted. Therefore scientists can use the ideas developed by other scientists immediately. The usual expectation is that the creator is cited, and indeed being cited by others is often what scholars most desire.

Patents provide another avenue for intellectual property to both encourage and restrain scientific and technological innovation. Patents provide protection for inventions for a limited time, which allows inventors to benefit from their creations, but can also restrain research and development. Sometimes companies buy patents covering inventions they never intend to use, as a means of suppressing competition with the technology that is currently the basis for their business. In other words, a rival technology is available, but it could threaten profits from current investments, so the rival technology is put on ice by purchasing patents covering it but not using them (Dunford, 1987). For example, General Electric obtained patents in order to slow down the introduction of fluorescent lights, in order to protect its sales of incandescent lights. In this case, a law designed to stimulate invention was used to suppress invention. Patent law is based on the presumption, sometimes false, that patent protection will be used to innovate rather than suppress innovation.

In the pharmaceutical industry, patents enable extraordinary profits, in conjunction with other techniques, especially marketing. Companies research new drugs, looking especially at ones that address chronic conditions such as arthritis, high cholesterol, and high blood pressure. These are especially profitable because drug use continues for months or years. When government regulators approve a suitable drug, massive marketing can establish it as a standard prescription. This marketing includes advertising, free samples for doctors, and free "educational trips" for doctors to seminars and conferences. Companies write papers about the drugs based on their in-house research and recruit academics to be ghost authors. Published in

leading medical journals, such papers give credibility to claims about the drug, and the company's marketing machine uses the publications as publicity.

When a drug patent expires, sometimes a company can patent and introduce a new drug that is quite similar and market it as superior. This is one of the methods of evergreening, which basically means using tricks to unfairly extend patent protection.

Because the criteria for granting patents are so easy to satisfy, some companies obtain dozens of patents for various aspects of a product, thereby preventing competitors from marketing competing products. By means of such "patent thickets," innovation is hindered. Then, there are patent trolls: companies that obtain patents and, rather than using them, search for companies that have inadvertently violated them, aggressively seeking payment in compensation. Patent trolling in essence uses intellectual property as a tool of extortion rather than innovation.

Pharmaceutical prescription drugs have become a source of major corruption (Gøtzsche, 2013). In some cases, researchers fiddle results, for example, by looking for adverse effects for only a short period or excluding certain types of people from studies (Goldacre, 2012). Drugs are promoted despite evidence that they are killing people. Some companies have been fined billions of dollars, indicating the massive scale of the corruption involved.

Intellectual property law is part of what enables abuses in the pharmaceutical industry. However, more significant than corruption is the distorting effect on research of the massive profits enabled by patent protection. The counterpoint to investigation of drugs that can be patented is a lack of investigation of substances that cannot be patented and indeed of anything that cannot be patented. For example, in one study, exercise was found to be as effective as antidepressants in dealing with depression (Craft & Perna, 2004). However, exercise cannot be patented, so companies have little incentive to study its benefits. The result is that billions of dollars are invested in researching and promoting antidepressants, while nonpatentable options, including diet, mindfulness, and exercise, are underresearched. The same sort of distortion of research agendas occurs in other areas.

In the case of pharmaceutical drugs, the impact of law on scientific advance is indirect, unlike the use of defamation law. Patent law offers a set of incentives that, in the hands of a powerful industry, compliant governments, and a willing medical profession, encourage research in some areas and starve it in others.

## Responses

When law operates to suppress or inhibit scientific research or findings, there are several possible responses. Three are described here: acquiescence, law reform, and resistance.

One option is to acquiesce by avoiding doing anything that might cause offence. In the case of defamation, this might mean not undertaking research that might trigger threats of legal action. Pieter Cohen, for example, could acquiesce by discontinuing his research into nutritional supplements. This

option is basically capitulation to legal obstacles. Put starkly, this might seem unacceptable to anyone committed to free inquiry. Yet, it is actually quite common, representing the chilling effect of the possibility of being sued. Many scientists avoid topics that might lead to adverse actions against them (public denunciations, loss of funding, threats of dismissal) and instead choose topics where there is ample funding and the promise of a welcoming response from others in the field.

A second response is to push for reform of laws that serve to inhibit research. Defamation law reform is an example. Michael Curtis (1995), in an article stimulated by Hilary Koprowski's defamation suit against *Rolling Stone* and Tom Curtis, argues for heightened legal protection for scientific speech. Similarly, Kate Sutherland (2010), in a discussion of Canadian defamation law inspired by a legal action against the publisher of a book review, argues for defamation law reform. Many commentators have recommended changes in laws on intellectual property (Halbert, 1999; Shulman, 1999). However, despite critiques of legal regimes and calls for law reform, in practice this path is both uncertain and slow. Concerns about scientific advance are a low priority in defamation law reform, where the interests of mass media and internet corporations are more influential, and in reform of intellectual property law, where the influence of the corporate beneficiaries of current law (software companies, pharmaceutical manufacturers, Hollywood producers, genetic engineering companies) is overwhelming.

A third possible response is to challenge the legal action by exposing it and mobilizing support against it, in an attempt to make the action counterproductive for the plaintiff. This approach is based on a model of outrage management, also called the backfire model (Martin, 2007). When a powerful individual or group does something that others might see as unfair – for example, sexual harassment, police beatings, massacres of peaceful protest, and genocide – the perpetrator often uses one or more methods to reduce public outrage:

- Cover up the action
- Devalue the target
- Reinterpret what happened by lying, minimizing consequences, blaming others, and reframing
- Use official channels to give an appearance of justice
- Intimidate or reward people involved

The classic case involving defamation is called McLibel. In the late 1980s, members of an anarchist group called London Greenpeace (not related to Greenpeace International) produced a leaflet titled "What's wrong with McDonald's?" telling, among other things, about poor working conditions for McDonald's workers and the unhealthy nature of McDonald's food. McDonald's, notoriously litigious, infiltrated the small London Greenpeace group, collected information, and sued five activists for defamation. Two of them, Helen Steel and Dave Morris, defended in court, triggering the formation of a large-scale support network. After the longest case in British history, McDonald's won in court but its reputation was severely damaged: it was a public relations disaster. McDonald's defamation action backfired (Donson, 2000; Vidal, 1997).

McDonald's used all five methods to reduce outrage. It tried to hide its use of infiltrators, devalued the members of London Greenpeace, reframed its action as defending the reputation of McDonald's, and used official channels (a court action) to make its action seem legitimate. The legal action intimidated three of the London Greenpeace activists, who capitulated.

Steel, Morris, and their supporters countered each one of these methods. They reproduced hundreds of thousands of copies of the leaflet "What's wrong with McDonald's?" and publicized the defamation action. Steel and Morris behaved with restraint. As ordinary workers (a gardener and a postman), they were hard to devalue. McLibel campaigners framed the defamation action as censorship. Campaigning itself went outside the legal domain. Finally, Steel and Morris resisted the intimidation of the legal action and refused to accept a generous settlement payment.

In summary, to counter the usual methods used to reduce outrage, opponents can increase outrage in these ways:

- Expose the action
- Validate the targets
- Interpret the events as unjust (censorship in the case of McLibel)
- Avoid or discredit official channels; instead, mobilize support
- Resist intimidation and rewards

This provides a general approach to addressing many uses of the law that inhibit scientific research that serves the public interest. In relation to defamation, the idea is to make legal threats and actions backfire by giving more attention to whatever is targeted for censorship (Jansen & Martin, 2003, 2015). The same approach can be used to analyze the struggle between the music industry and individuals who download songs, a case involving intellectual property (Martin, Moore, & Salter, 2010).

In the 1990s, the government of South Africa, to deal with the large number of AIDS cases, sought to import a generic HIV/AIDS drug: compulsory licensing and parallel importation are permitted by international intellectual property agreements. Nevertheless, dozens of pharmaceutical companies sued the government, putting their profits above the health of South Africans with AIDS. To challenge this abuse of power, AIDS activists and public health groups publicized the pharmaceutical companies' legal action, put the focus on AIDS patients, reframed the issue from patent law to health being sacrificed to corporate greed, mobilized support, and mounted numerous protest actions. The opponents of the companies thus used all five methods of increasing outrage (Halbert, 2005: 87–111).

## Conclusion

Three types of legal restraint on research were examined here. The first is the most obvious: defamation threats and actions against researchers that deter research on topics that would potentially benefit the community. There are relatively few cases

like this. However, defamation law has a more significant influence on research via the chilling effect: researchers will shy away from some topics because of the risk of being sued.

In the case of defamation, a key factor in restraining research is domain shifting. Rather than respond to research findings by scientific criticism or presenting contrary findings – the usual method in science – defamation suits shift the engagement to the legal domain, where money and legal technicalities take priority over scientific claims.

Laws against voluntary euthanasia are not directed at research but nonetheless discourage research on euthanasia by making research more difficult to undertake and by restricting funding for it.

Intellectual property provides a different sort of effect on research, via incentives. Patents, in conjunction with marketing and government regulatory processes, allow pharmaceutical companies to make massive profits from blockbuster drugs, thereby providing an incentive to prioritize investigating drugs with this potential. The spinoff effect is that research into other ways of improving health, including via exercise, diet, and nonpatentable substances, receives less attention than it would otherwise.

The implication of these case studies is that studying the adverse effects of law on science requires going beyond the most obvious cases of suppression. It is important to recognize that some restraints on research can be justified, so it is necessary to carefully assess the justifications. It is also important to look for indirect effects of laws, which can be deeper and more pervasive than the relatively few cases that receive attention.

**Acknowledgments** The author thanks David Allen, Pieter Cohen, Debora Halbert, Anneleis Humphries, and Qinqing Xu for useful feedback on drafts.

## References

Bagley, N., Carroll, A. E., & Cohen, P. A. (2017). Scientific trials—In the laboratories, not the courts. *JAMA Internal Medicine*, November 6. https://doi.org/10.1001/jamainternmed.2017.5730

Carroll, A. E. (2017). Why a lot of important research is not being done: Lawsuits have an intimidating effect on an already difficult enterprise. *New York Times*, December 4.

Côté, R. N. (2012). *In search of gentle death: The fight for your right to die with dignity*. Mt. Pleasant, SC: Corinthian Books.

Craft, L. L., & Perna, F. M. (2004). The benefits of exercise for the clinically depressed. *Primary Care Companion to the Journal of Clinical Psychiatry, 6*(3), 104–111.

Curtis, M. K. (1995). Monkey trials: Science, defamation, and the suppression of dissent. *William & Mary Bill of Rights Journal, 4*(2), 507–593.

Curtis, T. (1992). The origin of AIDS. *Rolling Stone* 626 (19 March), 54–61, 106, 108.

Drahos, P. (1996). *A philosophy of intellectual property*. Aldershot, UK: Dartmouth.

Donson, F. J. L. (2000). *Legal intimidation: A SLAPP in the face of democracy*. London, UK: Free Association Books.

Dunford, R. (1987). The suppression of technology as a strategy for controlling resource dependence. *Administrative Science Quarterly, 32*, 512–525.
Goldacre, B. (2012). *Bad pharma: How drug companies mislead doctors and harm patients*. London, UK: Fourth Estate.
Gøtzsche, P. C. (2013). *Deadly medicines and organised crime: How big pharma has corrupted healthcare*. London, UK: Radcliffe.
Hager, M. (2015). Confidentiality agreement handcuffs prominent assisted-suicide researcher. *The Globe and Mail*, January 6.
Halbert, D. J. (1999). *Intellectual property in the information age: The politics of expanding ownership rights*. Westport, CT: Quorum.
Halbert, D. J. (2005). *Resisting intellectual property*. London, UK: Routledge.
Jansen, S. C., & Martin, B. (2003). Making censorship backfire. *Counterpoise, 7*(3), 5–15.
Jansen, S. C., & Martin, B. (2015). The Streisand effect and censorship backfire. *International Journal of Communication, 9*, 656–671.
Katz, M. H., & Redberg, R. F. (2017). Science requires open discourse. *JAMA Internal Medicine*, November 6. https://doi.org/10.1001/jamainternmed.2017.5763
Magnusson, R. S. (2002). *Angels of death: Exploring the euthanasia underground. Melbourne*. Melbourne, VIC: University Press.
Martin, B. (2007). *Justice ignited: The dynamics of backfire*. Lanham, MD: Rowman & Littlefield.
Martin, B. (2010). How to attack a scientific theory and get away with it (usually): The attempt to destroy an origin-of-AIDS hypothesis. *Science as Culture, 19*(2), 215–239.
Martin, B., Moore, C., & Salter, C. (2010). Sharing music files: Tactics of a challenge to the industry. *First Monday, 15*(12).
Montgomery, B. (1994). Thesis claims "defame." *The Australian,* June 1, p. 21.
Ogden, R. D. (2010). Observation of two suicides by helium inhalation in a prefilled environment. *American Journal of Forensic Medicine and Pathology, 31*(2), 156–161.
Pring, G., & Canan, P. (1996). *SLAPPs: Getting sued for speaking out*. Philadelphia: Temple University Press.
Robbins, R. (2017). A supplement maker tried to silence this Harvard doctor — and put academic freedom on trial. *StatNews,* January 10.
Sheldrick, B. (2014). *Blocking public participation: The use of strategic litigation to silence political expression*. Waterloo, ON: Wilfred Laurier University Press.
Shulman, S. (1999). *Owning the future*. Boston: Houghton Mifflin.
Stark, D. E., & Shah, N. H. (2017). Funding and publication of research on gun violence and other leading causes of death. *Journal of the American Medical Association, 317*(1), 84–86.
Sutherland, K. (2010). Books reviews, the common law tort of defamation, and the suppression of scholarly debate. *German Law Journal, 11*(6), 656–670.
Vidal, J. (1997). *McLibel*. London, UK: Macmillan.

mitting the government to seize the assets of criminal organizations. The DOJ filed 1400 pages of evidence of misconduct on the part of the tobacco manufacturers who had engaged in a decades-long conspiracy to:

1. Mislead the public about the risks of smoking
2. Mislead the public about the danger of second-hand smoke
3. Misrepresent the addictiveness of nicotine
4. Manipulate the nicotine delivery of cigarettes to stimulate addiction
5. Market cigarettes misleadingly characterized as "light" or "low tar," while knowing that those cigarettes were at least as hazardous as full-flavored cigarettes
6. Target young smokers to ensure lifelong dependency
7. Reject the production of safer cigarettes, i.e., products with lower levels of nicotine (PHLC, 2010; TCLC, 2006)

In 2006, Judge Kessler issued a 1683-page opinion that found on the evidence that the tobacco companies had violated civil racketeering laws by lying for decades about the health risks of smoking and marketing to children. The DOJ sought to punish the companies by seizing assets obtained by this misconduct. However, the appeal court denied the government's remedy of a disgorgement of profits of $280 Billion (California HDE, 2005).The evidence suggested that the tobacco industry funded extensive pseudoscientific research in an attempt to discredit the efforts of various regulatory agencies to document the effects of environmental tobacco smoke, including second-hand smoke (Muggli et al., 2001).

In the 2006 decision Judge Kessler found that *"each and every one of these defendants repeatedly, consistently, vigorously - and falsely - denied the existence of any adverse health effects from smoking, despite the massive documentation in their internal corporate files from their own scientists, executives, and public relations people that confirmed that there was little evidence supporting their claims. Specifically, Defendants knew there was a consensus in the scientific community that smoking caused lung cancer and other diseases by at least January 1964. Despite this internal knowledge, the Defendants embarked on a campaign of proactive and reactive responses to scientific evidence that was designed to mislead the public about the health consequences of smoking"* (*US v. Philip Morris,* 2012). The court went on to say that the defendants publicly denied and distorted the truth about the addictive nature of nicotine, and designed their cigarettes to deliver the nicotine "sufficient to create and sustain addiction." The remedies consisted of an order issued in 2006 to publish "corrective statements" in advertisements on television, in newspapers, on the companies' websites and on cigarette packages to describe how the companies had misled the public. A preliminary agreement on how this was to be done was reached in October 2017, eleven years after the initial order was issued (Campaign TFK, 2017). The industry continues to face individual lawsuits from persons who have been affected by lung cancer and/or other tobacco-related diseases. In Canada, the provinces are negotiating with tobacco manufacturers to seek relief from costs inflicted on provincial health schemes from illnesses related to tobacco use. But tobacco remains legal and none of the tobacco executives who had the *mens rea* for decades have faced any criminal liabilities. Even after

being directed by the court during the *Philip Morris* trial to preserve all business records, 11 tobacco executives were found to have erased incriminating emails covering a two-and-a-half -year period prior to the initial verdict. The companies were fined $2.75 Million (Levin, 2004). Not the individuals.

## Beyond Tobacco: Exxon, Global Warming, and "Agnotology"

In 2015, a report appeared in *Scientific American* that expressly drew a parallel between Exxon and its knowledge of climate change, and the earlier history of tobacco. "Exxon was aware of climate change, as early as 1977, 11 years before it became a public issue . . . This knowledge did not prevent the company (now ExxonMobil and the world's largest oil and gas company) from spending decades refusing to publicly acknowledge climate change and even promoting climate disinformation—an approach many have likened to the lies spread by the tobacco industry regarding the health risks of smoking" (Hall, 2015). The journalists of the primary investigation of the Exxon case at *Inside Climate News* painted a more nuanced picture. In 1977, James F. Black gave a talk to senior executives suggesting that the expanding utilization of fossil fuels could lead to significant increases in greenhouse gases that would begin to warm the earth's atmosphere significantly (Banerjee, Song & Hasemyer, 2015). Within 2 years, the company's research division had commissioned a tanker, the Esso Atlantic, to measure the rate at which the oceans were absorbing $CO^2$, which it did from 1979 to 1982. Exxon also employed a team of mathematicians to prepare estimates of climate change based on complex atmospheric models. The work of Exxon scientists was published in various refereed journals between 1983 and 1984, and thereafter. Exxon was the sole leading oil and gas producer to take climate change seriously, and to develop an expertise in climate science.

Other scientists at Exxon warned of the development of an enormous natural gas find off Indonesia. It contained 70% $CO^2$ and would become the single largest source of $CO^2$ release on the globe if developed; it was not (Goldenberg, 2015). However, when the international community advocated the first steps to reduce carbon consumption by an international treaty at the Kyoto Summit, the chairman of Exxon, Lee Raymond, opposed it. For the next eleven years, Exxon funded climate change skeptics. In 2008, under mounting pressure from activist stakeholders, the company announced that it would end support for . . .[the] dozens of organizations who were actively distorting the science" (Banerjee et al., 2015). Currently, the Attorney General of New York has taken legal action to obtain corporate documents to determine if the company undertook a campaign to mislead shareholders and the public about global warming (Flitter, 2017). A 2017 study of company documents presented a rather ambiguous case against ExxonMobil based on a comparison of the publications of its scientists and the internal documents of executive versus what it suggested in its "advertorials" in the New York Times. "We conclude that ExxonMobil contributed to advancing climate science—by way of its scientists'

publications—but promoted doubt about it in its advertorials . . . We stress that the question is not whether ExxonMobil 'suppressed climate change research.' But rather how they communicated about it" (Supran and Oreskes, 2017).

The analogy between the tobacco case and the $CO^2$ case is not altogether convincing. Oreskes and Conway (2008, 2010) argue as though the "facts" behind climate change are completely incontrovertible and that there was a scientific consensus about them from the late 1970s. However, in a symposium on *Merchants of Doubt* (*Metascience,* 2012), scholars highly supportive of the research pointed out that it depicted science, particularly climate science, in a fashion that was inconsistent with studies of the actual practices of scientists in Science and Technology Studies, which emphasize the contingency, the boot-strapping logic, and idiosyncrasies of the discovery process. As Steve Yearly observes, "Oreskes and Conway are keen to emphasize the similarities between the work on these environmental and health topics and regular academic science . . . one cannot be a skeptic about the heliocentric solar system because the science is settled" (Yearly, 2012, 535) – implying that climate science is certainly *not* as settled as Newtonian physics. Yearly also points out that there has been a move away from science considered as an autonomous institution devoted to basic discovery to its increasing assignment in the post-WW2 state to enlarging the productivity of the economy, the military and medicine. And in the area of public health science, there is an increasing emphasis on risk assessment which necessarily involves public and political involvement in the regulatory process.

Assessing an optimum level for pesticide exposure, disposal of hazardous materials, etc. requires an estimation of *probable* safety levels, *probable* consequences and an evaluation of alternative solutions. These solutions "have to be offered in public forums where various interest groups have a legitimate role and where (the threat of) legal review is likely to be invoked" (p. 534).

David Mercer (2012, 537) argues in a similar vein. There is a tendency for "Oreskes and Conway's analysis to treat the boundaries between science, policy and regulation as clear and distinct," but in a democracy, where science is only possible by massive public investment, this is not the case. Furthermore, health science inevitably comes to play a role in governance, even though the science is not always "settled." The recent US report of global warming (CSSR, 2017) emphasizes that it has to develop policies based on two separate parameters: the *confidence* in the likelihood of change and the *impact* of the change should it occur. This approach recognizes the uncertainty of the measures and predictions, but unlike the tobacco "sound science movement" (Ong and Glantz, 2001), it does not freeze the regulatory agenda. In the case of global warming, the consequences of getting the policy wrong may prove to be catastrophic.

To return to the comparison with the tobacco case, a final point should be raised. "Sound science" counseled against regulation before the science was settled, but the advocates in the tobacco industry played a key role in creating the doubt. That was the rationale for promoting the term. And in the course of doing so, they lied to the public while millions of people died from the normal use of their products. To what extent is the charge comparable in the case of Exxon? To what extent had Exxon undermined effective public policies to protect the environment through its secrecy

and misrepresentations to the public? Or, on the contrary, to what extent have decisions about public policies been hobbled by technical incompleteness, debates about data manipulation, and the slow process of accumulating observations over the last few years as the current consensus has emerged, and as the international coalitions were proposed and adopted? At this point, no one can say with certainty. The exposé of tobacco is based on the disclosure of millions of pages of internal incriminating documents. No comparable record exists for Exxon.

There was another insidious aspect of the hold of tobacco on politicians and the media that differentiates it from the Exxon case: it stifled free speech. When *"60 Minutes"* produced a program on tobacco culpability and industry conspiracy, the program was spiked. When Stanton Glantz published the leaked tobacco papers on the website of USF, a congressional subcommittee took the unprecedented step of de-funding his studies of tobacco and health. And when Sharon Eubanks was successfully leading a RICO investigation against Philip Morris, persons associated with the Bush Presidency tried to undermine her prosecution. Tobacco lobbyists and lawyers were behind all of these cases. In a republic predicated on free speech, the power of corporate actors to suppress criticism is injurious to the free exchange of ideas and, in this case, the negotiation of effective policies to protect public health.

We do not have to draw any conclusions about Exxon at this point, but there is a more general lesson. It is raised through the term, "agnotology," coined by Robert Proctor (Proctor & Scheibinger, 2008). Recalling Nietzsche, it might be called *the genealogy of ignorance*. Often, the absence of knowledge is not a natural condition of society, but an outcome of concerted, institutional efforts to suppress knowledge, sow confusion, disappear the past, suppress unwanted voices, and occlude competing world views. In this essay, we have attempted to enlarge the study of groupthink – which emphasizes how people come to give erroneous accounts of the world – to conditions where knowledge of reality is actively and institutionally suppressed or distorted. Tobacco "science" represents a compelling case study in agnotology.

## References

Banerjee, N., Song, L., & Hasemyer, D. (2015). The road not taken: Exxon's own research confirmed fossil fuels' role in global warming decades ago. *Inside Climate News* online, https://insideclimatenews.org/content/Exxon-The-Road-Not-Taken. Retrieved 26 July 2019.

California HDE. (2005). *California healthline daily edition "supreme court refuses to hear appeal on past profits in doj racketeering case against tobacco companies"*, Online https://californiahealthline.org/morning-breakout/supreme-court-refusesto-hear-appeal-on-past-profits-in-doj-racketeering-case-against-tobaccocompanies/. Retrieved 10 Nov 2017.

Campaign TFK. (2017). Tobacco companies must finally tell public the truth about their lethal product – 11 years after a court ordered it. *Press Release* on October 4, https://www.tobaccofreekids.org/press-releases/2017_10_04_corrective. Retrieved 4 Nov 2017

CSSR. (2017). *Climate science special report: Fourth national climate assessment*. Washington, D.C.: US Global Change Research Program. https://doi.org/10.7930/J0J964J6

Flitter, E. (2017, June 2). New York prosecutor says Exxon misled investors on climate change. *Reuters Business News*. https://www.reuters.com/article/ususa-climatechange-exxon/new-york-prosecutor-says-exxon-misled-investors-onclimate-change-idUSKBN18T1XK Retrieved 4 Nov 2017.

Glantz, S., Slade, J., Bero, L. P., Hanauer, P., & Barnes, D. (1996). *The cigarette papers*. Berkeley, CA: University of California Press.

Goldenberg, S. (2015, July 9) Exxon knew of climate change in 1981, email says – But it funded deniers for 27 more years. *The Guardian* online, https://www.theguardian.com/environment/2015/jul/08/exxon-climate-change-1981-climate-denier-funding. Retrieved 26 July 2019.

Hall, S. (2015). Exxon knew about climate change almost 40 yeas ago. *Scientific American*, Online October 26, https://www.scientificamerican.com/article/exxon-knew-about-climate-change-almost-40-years-ago/. Retrieved 26 July 2019.

Levin, M (2004, July 22). Judge orders Philip Morris, Altria to pay $2.75 Million. *Los Angeles Times*. http://articles.latimes.com/2004/jul/22/business/fi-smoke22. Retrieved 4 Nov 2017.

Mercer, D. (2012). Symposium comments "perspectives on global warming". *Metascience, 21*, 235–240. https://doi.org/10.1007/s11016-011-9639-9

Metascience. (2012). Perspectives on global warming. *Australasian Association for the History, Philosophy and Social Studies of Science, 21*, 531–559. https://doi.org/10.1007/s11016-011-9639-9

Michaels, D. (2008a). *Doubt is their product*. New York, NY: Oxford University Press.

Michaels, D. (2008b). Manufactured uncertainty: Contested science and the protection of the public's health and environment. In R. N. Proctor & L. Schiebinger (Eds.), *Agnotology: The making and unmaking of ignorance* (pp. 90–107). Stanford, CA: Stanford University Press.

Muggli, M. E., Foster, J. L., Hurt, R. D., & Repace, J. L. (2001). The smoke you don't see: Uncovering tobacco industry scientific strategies aimed against environmental tobacco smoke policies. *American Journal of Public Health, 91*(9), 1419–1423.

Norr, R. (1952, December). Cancer by the carton. *Reader's Digest*, 7–8.

Ong, E. K., & Glantz, S. (2001). Constructing 'sound science' and 'good epidemiology': Tobacco, lawyers and public relations firms. *American Journal of Public Health, 91*(11), 1749–1757.

Oreskes, N., & Conway, E. M. (2008). Challenging science: how climate science became a victim of the cold war. In R. N. Proctor & L. Schiebinger (Eds.), *Agnotology* (pp. 55–89). Stanford, CA: Stanford University Press.

Oreskes, N., & Conway, E. M. (2010). *Merchants of doubt*. New York, NY: Bloomsbury Press.

PHLC. (2010). *United States v. Philip Morris (D.O.J. Lawsuit)*. St. Paul, MN: Public Health Law Center at William Mitchel College of Law. http://www.publichealthlawcenter.org/topics/tobacco-control/tobacco-controllitigation/united-states-v-philip-morris-doj-lawsuit/. Retrieved 3 Nov 2017.

Player, T. S. (1998). After the fall: The cigarette papers, the global settlement, and the future of tobacco litigation. *South Carolina Law Review, 49*, 311–342.

Proctor, R. N. (1999). *The nazi war on cancer*. Princeton, NJ: Princeton University Press.

Proctor, R. N. (2011). *Golden holocaust: origins of the cigarette catastrophe and the case for abolition*. Berkeley, CA: University of California Press.

Proctor, R. N., & Schiebinger, L. (Eds.). (2008). *Agnotology: the making and unmaking of ignorance*. Stanford, CA: Stanford University Press.

Rabin, R. L. (1992). A Sociolegal history of tobacco tort legislation. *Stanford Law Review, 44*(4), 853–878.

Supran, G., & Oreskes, N. (2017, August 23). Assessing ExxonMobil's climate change communications (1977–2014). *Environmental Research Letters, 12*(8). http://iopscience.iop.org/article/10.1088/17489326/aa815f?fromSearchPage=true. Retrieved 5 Nov 2017.

TCLC. (2006). *The Verdict is in: Findings from United States v. Philip Morris*. Saint Paul, MN: Tobacco Control Legal Consortium.

U.S. (1964). *Surgeon general's report, Smoking and health: report of the advisory committee to the surgeon general of the public health service*. Washington, D.C.: USGPO.
U.S. (2014). *Surgeon general's report, The health consequences of smoking: 50 years of progress*. Washington, D.C.: USGPO.
U.S. v. Philip Morris. (2012). Judgment of United States District Court for the District of Columbia, Civil Action No. 99-2496 (GK), November 27.
Yearly, S. (2012). Symposium comments, "Perspectives on global warming". *Metascience, 21*, 531–535. https://doi.org/10.1007/s11016-011-9639-9.

# Chapter 14
# A Plea for Global Consideration of Human Brain Sex Differences


**James W. Howell**

## Introduction

A woman looking at medical science today can find herself in a perplexing situation. She may have heard the recent criticisms that medical and biological research is mostly done with men and not with women. She may have heard that the excuse for this was related to budget issues. There was a reluctance to spend the money and the time adjusting research to female cycles and differences in the anatomy and physiology of males and not females and of men and not women. Sexual dimorphism exists throughout the human body. Any individual patient wants to get a diagnosis and treatment that is proper for who they are and appropriate to their age and condition.

Confounding this issue is a movement within some groups in science questioning sexual dimorphism. Somehow the proponents of this movement have managed to particularly focus on the human brain, as if this body part in some way had no interaction with the other parts of the body and managed to evolve at its own separate pace and manner.

As you will see in this chapter, this way of thinking can put patients in dangerous situations. When you make a systematic study of the various organs of the body, as you will see in the brief descriptions of some parts of the human body in this chapter, it becomes readily apparent that there are vital differences between the anatomy and physiology of the woman's body and that of the man.

It is also very true that there have been many destructive and false ideas advanced over time about supposed biological differences between men and women that are not based on science at all but either on folklore or sexist ideas such as that women are "too emotional" to serve in an executive capacity. This does not mean, of course,

J. W. Howell (✉)
Department of Psychiatry, University of Tennessee Health Science Center, Memphis, TN, USA
e-mail: atenra@comcast.net



D. M. Allen, J. W. Howell (eds.), *Groupthink in Science*,
https://doi.org/10.1007/978-3-030-36822-7_14

that there are no *real* differences. The proponents of the movement to deny sexual dimorphism make up a dangerous groupthink force that attempts to stifle those who disagree with their doctrines by calling their opponents misogynists enemies who want to discriminate against women.

Other scientists such as Debra Soh try to make it clear that denying science will not in fact do anything to fight misogyny (Lehman and Soh, 2017). The misogynists, bigots, and people who wish to discriminate against women will always find ways to spread their hatred and regressive ideas. The denial of science and sexual dimorphism will just spread ignorance and put women in danger.

In her paper on sex differences in the cardiovascular systems of men and women, Virginia Huxley (2007) emphasized the importance of understanding the dimorphism of the two systems. This improves diagnostic systems, the recognition of sex-specific pathophysiology, and the development and implementation of proper treatment for each of the sexes. She emphasized the fundamental importance of realizing the fact that each cell in the body is either XX or XY from the time the organism is in the uterus, through prepuberty, to adulthood.

Margaret McCarthy, Arnold, Ball, Blaustein, and DeVries (2012) went further in discussing sex differences by presenting a description of nonexclusive categories that would help in developing experimental designs:

1. The first type is absolute sexual dimorphism. This includes two-component sets of particular behavioral, physiological, or morphological forms, one found in the male and one in the female. Copulatory behavior would be an example.
2. The second type exists along a continuum or sliding scale in which any given male or female can be found at any point, but the *average* of individuals would differ between the sexes. Odor detection and learning are examples of this.
3. The third type, and most complicated to understand, involves characteristics which might converge at some endpoint or diverge after some challenge. The neurophysiology that regulates one of these behaviors might be completely different in the male and the female. Sex-specific parental behavior could be an example and might manifest itself completely differently from one species to another.

In considering the effect of accepting the idea that there are sex differences, one particular assumption has had a deleterious effect. Too many of the criticisms of sexual dimorphism in humans are rooted in what McCarthy (2016) calls the pervasive assumption that "sex difference in neuroanatomy and neurophysiology is synonymous with a sex difference in behavior." Such an assumption in a particular case would have to be tested.

There is a growing body of literature concerning sexual dimorphism. Margaret McCarthy's papers on the subject are a great place to start familiarizing yourself with this literature, but other references include Shansky (2016), Plaff and Christen (2013). In 2015, the (NIH 2015) made it mandatory, because there are physiological and anatomical differences between the sexes, that all research use sex balanced cohorts and treat sex as a biological variable. This was reaffirmed in later years.

## Global Considerations

In Lise Eliot's review of Gina Rippon's book, The Gendered *Brain: The New Science That Shatters the Myth of the Female Brain*, she made a statement that seems to indicate a lack of understanding of anatomy, physiology, or evolution: "The brain is no more gendered than the liver, kidneys, or heart."

First, this ignores the fact that every cell of these organs has an XX or XY chromosome pair, marking the sex of the individual.

Second, according to an extensive literature, there are sex difference effects in most organs of the body. A few details about the liver, kidneys, and heart are considered below.

### *Sex Differences in Human Gut and the Brain*

The human gut is a particularly striking example of sexual dimorphism. The gut and the human brain work closely together. The gut has been implicated in contributing to intuitive decision making, affect, components of language, higher cognitive functions, motivation, emotion regulation, and gastrointestinal homeostasis. In addition, the intestinal microbes and host microbes work with the nervous system's interaction with the brain to form what many call the *enteric nervous system* (the ENS, sometimes referred to as the "second brain") (Mayer, 2011). There are pronounced differences in the dynamics of microbial growth and effects both over time and between men and women. This is true even when comparing diverse ethnic and widely separated cultural groups (de la Cuesta-Zuluaga et al., 2019).

Gut microbiota seem to regulate the synthesis and release of oxytocin, which has an effect on parturition and lactation.

### *Human Olfaction Sex Differences*

Sensitivity to smell varies according to sex among children (Schriever et al., 2018). Although most investigators have agreed since at least 1899 (Toulouse and Vaschide) that the abilities of women for olfaction are superior to men, some studies that involve large samples suggested the abilities between the sexes do not differ all that much. However, a meta-analysis of thousands of men and women in existing studies focused on sex differences in identification, discrimination, and threshold confirmed that women's olfactory abilities are greater than those of men (Sorokowski et al., 2019).

Doty and Cameron (2009) suggested that one possible explanation for this finding is interactions between early experiences of smell perception in certain brain regions with circulating endocrine substances. This, combined with later

hormonal mechanisms in an adult woman's life, could result in the superior olfactory perception. Another possible explanation is that men have lesser verbal skills than women making it easier for women to answer questions in the experimental process (Larsson, Finkel, & Pedersen, 2000, Oberg, Larsson, & Backman, 2002).

## *Renal Function Sex Differences*

In both mice and humans, persistent differential gene expression between the sexes in kidney function includes drug and steroid metabolism as well as osmotic regulation in a study by Rinn et al. (2004).

## *Sex Differences in the Cardiovascular System*

Cardiovascular data reflected in textbooks, handbooks, and relevant Internet sites usually come from 18- to 22-year-old healthy males. As mentioned, the reason given is that authors were avoiding the "confusing" problem of cycling that would have to be considered when including data from women. Women are found to have lower norepinephrine levels than men and a host of other differences of which medical professionals should be aware when treating pathologies related to the heart. Even when considering the three typical hallmarks of men's heart attacks as described in the medical literature, the fact is that only one in three women will experience these symptoms when they have a myocardial infarction. According to Virginia Huxley (2007) those hallmarks are:

1. Chest discomfort or uncomfortable pressure, fullness, squeezing, or pain in the center of the chest that lasts longer than a few minutes or that comes and goes.
2. Spreading pain to one or both arms, back, jaw, or stomach.
3. Cold sweats and nausea.

In fact, a woman having a heart attack may well have other symptoms such as vomiting or back or jaw pain. It is important that sex differences be recognized, included in medical training, and used to diagnose and treat disease (Huxley, 2007).

In 2010, John Konhilas published an extensive review of the literature in which he further discussed the differences men and women experience with heart disease, especially congestive heart failure (CHF).

### *Sex Differences in the Liver*

Krebs et al. (2003) describe how in the liver, as elsewhere, there is a complex interplay of hormonal, developmental, and tissue-specific control of gene expression. This leads to tissues which are found in two distinct forms in males versus females. For example, sex-specific patterns of liver gene expression occur in the production of several enzymes involved in the metabolism of steroids and as well as for the metabolism of synthetic chemicals. The extent and duration of the activation of certain hepatic genes are dependent on the nature of growth hormone signaling as well as interactions with numerous other proteins within the cells. Krebs adds that hepatic sex differences may prove relevant to medical issues that vary with gender, such as differences in drug metabolism and the incidence of certain diseases, as well as to problems related to pregnancy.

## Twin Studies

Although twin studies are not definitive because of the extreme difficulty of separating out purely genetic effects from gene-environment interactional factors, they clearly show that genetic differences (such as the presence or absence of a Y chromosome) can create differences in anatomy, physiology, the endocrine system, and behavioral tendencies (although not specific behaviors).

## Conclusions

The primary message of this chapter is that future work in the study of sex differences should include a broad investigation of as many aspects of the animal body as possible. The limited number of global considerations outlined here underlines the importance of doing this. Of course, there are many other organs, systems, and body functions that could have been included. Additionally, the few that have been included here have not been discussed exhaustively.

The brain is not isolated. Parts of the body are acting on it and the brain, of course, serves to regulate and maintain the body. This should be an obvious conclusion even after this brief glance of the literature.

Debra Soh said on March 11, 2019, in *Quillette* that, “Denying science won’t end sexism,” and that the people rejecting sexual dimorphism actually are questioning the value of feminism. This reminds me of a quote by one of the earliest feminist writers, Mary Wollstonecraft, who, when writing about her experiences in her book, *The French Revolution*, said, “Every political good carried to the extreme must be productive of evil (1790).”

## Bibliography


Arnold, A., & Breedlove, A. (1985). Organizational and activational effects of sex steroids on brain and behavior: A reanalysis. *Hormones and Behavior., 19*(4), 469–498. https://doi.org/10.1016/0018-506x(85)90043-X

Baron-Cohen, S., Knickmeyer, R., & Belmonte, M. (2005). Sex differences in the brain: Implications for explaining autism. *Science, 310*, 819–823.

Broere-Brown, Z., Baan, E., Schalekamp-Timmermans, S., Verburg, B., Jaddoe, V., & Steegers, E. (2016). Sex-specific differences in fetal and infant growth patterns: A prospective population-based cohort study. *Biol Sex Differ, 7*, 65.

de la Cuesta-Zuluaga, J., Kelley, S. T., Chen, Y., Escobar, J. S., Mueller, N. T., Ley, R. E., et al. (*2019*). Age- and sex dependent patterns of gut microbial diversity in human adults. *mSystems, 4*, e00261–e00219. https://doi.org/10.1128/mSystems.00261-19

Doty, R. L., & Cameron, E. L. (2009). Sex differences and reproductive hormone influences on human odor perception. *Physiology & Behavior, 97*, 213–228. https://doi.org/10.1016/j.physbeh

Greenberg, D., Warrier, V., Allison, C., & Baron-Cohen, S. (2018). Testing the empathizing-systemizing theory of sex differences and the extreme male brain theory of autism in half a million people. *PNAS, 115*(48), 12152–12157. https://doi.org/10.1073/pnas.1811032115

Huxley, V. H. (2007). Sex and the cardiovascular system: The intriguing tale of how women and men regulate cardiovascular function differently. *Advances in Physiology Education, 31*, 17–22. https://doi.org/10.1152/advan.00099.2006

Ingalhalikar, M., Smith, A., Parker, D., Satterthwaite, D., Elliott, M., Ruparel, K., et al. (2014). Sex differences in the structural connectome of the human brain. *Proceedings of the National Academy of Sciences, 111*(2), 823–828.

Jazin, E., & Cahill, L. (2010). Sex differences in molecular neuroscience: From fruit flies to humans. *Nature Reviews Neuroscience, 11*, 9–17.

Knickmeyer, R. C., Wang, J., Zhu, H., Geng, X., Woolson, S., Hamer, R. M., et al. (2014). Impact of sex and gonadal steroids on neonatal brain structure. *Cerebral Cortex, 24*(10), 2721–2731.

Konhilas, J. P. (2010). What we know and do not know about sex and cardiac disease. *Journal of Biomedicine and Biotechnology*, 562051. https://doi.org/10.1155/2010/562051

Kopsida, E., Stergiakouli, E., Lynn, P. M., Wilkinson, L. S., & Davies, W. The role of the Y chromosome in brain function. *The Open Neuroendocrinology Journal, 2*, 20–30. https://doi.org/10.2174/1876528900902010020

Krebs, C. J., Larkins, L. K., Price, R., Tullis, K. M., Miller, R. D., & Robins, D. M. (2003). Regulator of sex-limitation (Rsl) encodes a pair of KRAB zinc-finger genes that control sexually dimorphic liver gene expression. *Genes & Development, 17*(21), 2664–2674. https://doi.org/10.1101/gad.1135703

Lenroot, R. K., Gogtay, N., Greenstein, D. K., Wells, E. M., Wallace, G. L., Clasen, L. S., et al. (2007). Sexual dimorphism of brain developmental trajectories during childhood and adolescence. *NeuroImage, 36*(4), 1065–1073. https://doi.org/10.1016/j.neuroimage.2007.03.053

Leonard, C., Towler, S., Welcome, S., Halderman, L., Otto, R., Eckert, M., et al. (2008). Sex matters: Cerebral volume influences sex differences in neuroanatomy. *Cerebral Cortex, 18*(12), 2920–2931. https://doi.org/10.1093/cercor/bhn053

Luders, E., Gaser, C., Narr, K. L., & Toga, A. W. (2009, November 11). Why sex matters: Brain size independent differences in gray matter distributions between men and women. *The Journal of Neuroscience, 29*(45), 14265–14270.

Mayer, E. A. (2011, July 13). Gut feelings: The emerging biology of gut-brain communication. *Nature Reviews. Neuroscience, 12*, 453–466. https://doi.org/10.1038/nm3071

McCarthy, M. M. (2016). Multifaceted origins of sex differences in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*, 20150106. https://doi.org/10.1098/rstb.2015.0106

McCarthy, M. M., Arnold, A. A., Ball, G. F., Blaustein, J. D., & DeVries, G. J. (2012, February 15). Sex differences in the brain: The not so inconvenient truth. *The Journal of Neuroscience, 32*(7), 2241–2247.

NIH (2015). Consideration of sex as a biological variable in NIH-funded Research. Notice Number: NOT-OD-15-102.

O'Conner, C., & Joffe, H. (2014). Gender on the brain: A case study of science communication in the new media environment. *PLoS One, 9*(10), e110830. https://doi.org/10.1371/journal.pone.0110830

Plaff, D. W., & Christen, Y. (Eds.). (2013). *Multiple origins of sex differences in brain: Neuroendocrine functions and their pathologies*. New York, NY: Springer.

Polderman, T. J., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., et al. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics, 47*, 702. https://doi.org/10.1038/ng.3285

Rinn, J. L., Rozowsky, J. S., Laurenzi, I. J., Petersen, P. H., Zou, K., Zhong, W., et al. (June 2004). Major molecular differences between mammalian sexes are involved in drug metabolism and renal function. *Developmental Cell, 6*, 791–800.

Rippon, G. (2019). The Gendered Brain: The new neuroscience that shatters the myth of the female brain. New York, NY: Vintage Publishing.

Ruigrok, A., Salimi-Khorshidi, C., Lai, M., Baron-Cohen, S., Lombardo, M., Tait, R., et al. (2013). A meta-analysis of sex differences in human brain structure. *Neuroscience and Biobehavioral Reviews*. https://doi.org/10.1016/j.neubiorev.2013.12.004lumenthal

Savic, I. (Ed.). (2010). *Sex differences in the human brain, their underpinnings and implications*. Philadelphia, PA: Elsevier.

Schriever, V. A., Agosin, E., Altundag, A., Avni, H., Van, H. C., Cornejo, C., et al. (2018). Development of an international odor identification test for children: The Universal Sniff Test. *The Journal of Pediatrics, 198*, 265–272.e3. https://doi.org/10.1016/j.pes.2018.03.011

Shansky, R. M. (Ed.). (2016). *Sex differences in the central nervous system*. Philadelphia, PA: Elsevier.

Snell, D. M., & Turner, M. A. (2018). Sex chromosome effects on male-female differences in mammals. *Curr Biol, 28*(22), R1313–R1324. https://doi.org/10.1016/j.cub.2018.09.018

Soh, D. (2019). Science denial won't end sexism. Quillette

Sorokowski, P., Karwowski, M., Misiak, M., Marzak, M. K., Dziekan, M., Hummel, T., et al. (2019). Sex differences in human olfaction: A meta-analysis. *Frontiers in Psychology, 10*, 242. https://doi.org/10.3389/fpsyg.2019.0024

Toulouse, E., & Vaschide, N. (1899). Mesure de l'ordorat chez l'homme et chez la femme. *CompteRendus Social Biology*, 381–383.

Tyan, Y., Liao, J., Lin, Y., & Weng, J. (2017). Gender differences in the structural connectome of the teenaged brain revealed by generalized q-sampling MRI. *Science Digest*. https://doi.org/10.1016/j.ncil.2017.05.014

Wang, C. (2018). Decoding sex differences in the brain, one worm at a time. *Gender and the Genome, 2*(3), 76–80. https://doi.org/10.1177/2470289718789306

Wheelock, M. D., Hect, J. L., Hernandez-Andrade, E., Hassan, S. S., Eggebrect, A. T., & Thomason, M. E. (2019). Sex differences in functional connectivity during fetal brain development. *Developmental Cognitive Neuroscience, 36*, 100632.

# Chapter 15
# Ideological Blinders in the Study of Sex Differences in Participation in Science, Technology, Engineering, and Mathematics Fields


**David C. Geary and Gijsbert Stoet**

There is little question that there are sex differences in engagement in certain science, technology, mathematics, and engineering (STEM) fields. The U.S. National Science Foundation (NSF), for instance, reports that women are awarded 57% of all undergraduate STEM degrees (compared to 61% of non-STEM degrees) but with substantial differences across fields. Women earn the majority of degrees in the life and social sciences, but less than 20% of the degrees in computer science and engineering (http://www.nsf.gov/statistics/2015/nsf15311/tables.cfm). In other words, the sex differences in STEM degrees and in later occupational choices are largely in inorganic fields, those focused on understanding non-living things as contrasted with living things. These differences are practically important because they and more general differences in the type of occupations men and women enter contribute, in part, to the sex difference in earnings (Del Río & Alonso-Villar, 2015).

These sex differences and the social prestige of many STEM occupations have generated a cottage industry within academia, the popular media, and beyond. The movement is fueled by the zeitgeist among some feminist activists that there should be gender equality – equal *outcomes* regardless of any underlying sex differences in academic or occupational interests or in the patterns of cognitive strengths – for anything of monetary or social value. In this case, the focus is on identifying and eliminating the causes of the STEM discrepancies (e.g., Hill, Corbett, & St Rose, 2010). As an example of the resources devoted to achieving equality, since 2001 the NSF has invested more than $130 million into the ADVANCE program (Advancement

D. C. Geary (✉)
Department of Psychological Sciences, Interdisciplinary Neuroscience Program, University of Missouri, Columbia, MO, USA
e-mail: GearyD@missouri.edu

G. Stoet
Department of Psychology, University of Essex, Colchester, Essex, UK
e-mail: g.stoet@essex.ac.uk



D. M. Allen, J. W. Howell (eds.), *Groupthink in Science*,
https://doi.org/10.1007/978-3-030-36822-7_15

of Women in Academic Science and Engineering Careers: http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5383) in an attempt to close the gap in STEM disciplines with similar efforts instituted in other Western countries (e.g., http://www.ecu.ac.uk/equality-charters/athena-swan/). Many of the activities funded by these initiatives make sense and are likely to be helpful in some ways, such as developing mentoring programs for women who are junior faculty in science and engineering in university settings although it raises ethical questions when the same mentoring programs are not provided for male junior faculty, as is case in the UK's Athena SWAN Swan's programs. There are, in addition, other themes regarding the sources of these differences that are based on weak evidence and a large dose of wishful thinking. The most questionable and perhaps the most favored of these are stereotype threat, implicit bias, and microaggression.

Stereotype threat allegedly occurs when one is confronted with tasks or situations that trigger negative stereotypes (e.g., that 'women are not as proficient at math as men') that in turn results in a preoccupation about performing in a way that confirms the stereotype (Spencer, Steele, & Quinn, 1999). Critically, the preoccupation is said to undermine actual performance even when there is no factual basis to the stereotype. Implicit bias is a related concept and involves an unconscious association between group membership (e.g., sex or race) and stereotypical positive or negative attributes that in turn can result, in theory, in prejudicial behavior toward individuals within that group (Greenwald, McGhee, & Schwartz, 1998; Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Microaggressions are subtle behaviors (e.g., facial expressions) or statements that are not explicitly hostile but are nevertheless interpreted by the receiver as conveying contempt, stereotypical attitudes, or other negative beliefs. Examples of verbal microaggressions are provided by the University of California, Santa Cruz (e.g., 'You're a girl, you don't have to be good at math', https://academicaffairs.ucsc.edu/events/documents/Microaggressions_Examples_Arial_2014_11_12.pdf).

The basic argument is that some significant proportion of the sex differences in STEM fields – but only those in which men outnumber women – is thought to be caused by pervasive negative stereotypes about women's abilities in these fields that in turn undermine their performance. And, by poor treatment by STEM teachers and colleagues – microaggressions – that seeps from their unconscious belief in these same stereotypes to create unsupportive and even subtly hostile classrooms and work environments. These types of explanations fit well with the narrative of some gender activists: that the sex difference is largely due to social and cultural factors that undermine women's pursuit of degrees and occupations in STEM fields (Hill et al., 2010).

In any case, these concepts have been embraced by the mass media and beyond. Examples of this embrace include accusations in the *New York Times* that the wording of several SAT items will trigger stereotype threat and undermine girls' performance on the mathematics section of the test (Hartocollis, 2016) and self-help books to cope with one's own unconscious biases (Thiederman, 2015). On the face of it, there is nothing wrong with academic and mass media focus on these topics, as related to sex differences in STEM participation. The real issues concern the magnitudes of these effects on women's STEM participation and the foregone

opportunities of not focusing on other factors that might have an even stronger impact on their participation.

Let us consider first the magnitude of stereotype threat on girls' and women's mathematics achievement. As noted, the concept is now widely known in popular culture and the first scientific publication on the topic has been cited more than 3000 times in Google scholar (Spencer et al., 1999), a seminal contribution by this measure. Accordingly, it is not surprising that there are now interventions to counter the hypothesized negative effects of stereotype threat on women's performance in STEM fields (e.g., Walton, Logel, Peach, Spencer, & Zanna, 2015). Given the prominence of the topic and the resources devoted to it, we carried out the first meta-analysis (i.e., statistical aggregation of experimental results across many studies) of the effect of stereotype threat on sex differences in mathematics performance (Stoet & Geary, 2012). We reasoned that if stereotype threat had a substantive effect on girls' and women's mathematics performance then the most basic experimental manipulation of the effect should replicate across studies.

The design is simple and includes four groups: one group of women and one group of men who take a mathematics test under typical testing conditions (control group), and groups that take the test under threat conditions (experimental group). The latter might involve telling participants that men typically do better on the mathematics test. In theory, men in the experimental and control conditions should perform about the same on the test, but women in the threat condition should perform worse than women in the control condition. One would think that there would be hundreds of studies that have used this basic design, but most of the replications in this field (social psychology) are 'conceptual' and not exact; conceptual is based on creating conditions that should replicate the basic idea (that threat will compromise women's performance) rather than replicate the exact experimental procedures. We found 20 studies that were very similar to the basic experimental design followed by Spencer et al. (1999), and only 11 of them replicated their effect. Of the 11 that found an effect, only 3 did not rely on a controversial statistical control that might exaggerate any such effect.

We could not definitively conclude from our analyses that stereotype threat does not exist, but we did question whether the magnitude of any such effect merited the scientific and popular press attention it was receiving. This of course is not likely to be a popular conclusion, based on the above-described interest in the phenomenon, and indeed it was not. We sent the manuscript to three or four journals before an editor would even send it for peer review, a pattern that we have found for nearly all of our subsequent sex differences studies that reached unpopular conclusions; one of us (Geary) has the same experiences in his work on biological sex differences and the other of us (Stoet) has the same experience in his other work on educational sex differences. In this case, one of these is a very prominent journal in the field of psychology and the editors took three months – and this was only after several inquiries regarding the status of the submission – before they informed us that it would not be sent for peer review, indicating that failures to replicate (follow-up experiments that cannot confirm an original finding) were not of interest to them; this was before the emergence of the replication crisis in social psychology and the attendant focus on

replications. Editors rejecting manuscripts without peer review are common but this is typically done within one or at most two weeks, not three months. After the article was published, we were greeted by an angry response by several proponents of stereotype threat, not the dispassionate curiosity as to why the effect is sometimes found and sometimes not.

In a related analysis, Flore and Wicherts (2015) found a similar overall (small) effect, but when they corrected for publication bias – the tendency for positive but not negative results to be published – the effect essentially disappeared. This means that there is evidence for a small stereotype threat effect in the scientific literature, but because studies that do not find an effect tend not to get published in this literature, the real-world impact of stereotype threat is probably close to zero (see also Ganley et al., 2013; Picho, Rodriguez, & Finnie, 2013). Picho et al. (2013) also found evidence for publication bias but discounted its importance. At the time of the writing of this chapter, a large replication effort is being carried out, and we are optimistic that this and other similar research focusing on replicability can give a definite answer on the question of whether stereotype threat can undermine girl's and women's performance in mathematics and if so, determine the magnitude of this effect. It should be noted, though, that the largest study carried out thus far with nearly 1000 students found no effects (Ganley et al., 2013). This latter study is of particular relevance, because it was carried out with adolescents and school children. If stereotype threat discourages girls from pursuing math-intensive STEM coursework and careers, its effect should be evident in adolescence. The fact that a large and well-designed study could not find any effect, in our opinion, suggests either the effect does not exist or it is unmeasurably small.

Either way, the existing evidence indicates that stereotype threat has received outsized attention from educational policy makers and opinion makers. The bottom line is that there is at best a small and probably no effect at all of stereotype threat on women's mathematical performance. Thus, the considerable efforts at addressing this 'problem' will almost certainly have little if any effect on girl's and women's participation in inorganic STEM fields.

We suspect the same is true for implicit bias. For a variety of cultural and legal reasons, the level of explicit sexism has dropped considerably over the years in most school and work environments. But, girls' interest and women's participation in inorganic STEM fields has remained stubbornly low over the past 20 years (Hill et al., 2010). So, there are two options. One might conclude that explicit sexism is no longer keeping girls and women away from these fields and so something else must be contributing to these sex differences. Or, one can maintain the conceptual grasp on sexism as a causal factor and switch focus to an 'unconscious' subtle form of sexism that results from implicit bias (see Greenwald et al., 1998; Greenwald et al., 2009) and its behavioral companion, microaggression (Basford, Offermann, & Behrend, 2014).

Indeed, implicit bias has achieved a cult-like status in some academic circles and in the wider culture. There are now on-line tests to assess one's implicit bias in a number of areas, including sex differences in work and family. We are not doubting that people do have all sorts of implicit beliefs that may or may not be accurate. The issues here are whether we can rigorously and accurately assess these biases, and

whether the strength of any such biases is sufficient to explain the sex differences in STEM fields. The assessment of implicit bias is often done using the implicit associations test (e.g., https://implicit.harvard.edu/implicit/user/agg/blindspot/indexgc.htm) whereby the strength of people's associations between sex (or race) and certain attributes, such as work or science, is assessed by a series of categorization tasks. The difference between the speed of categorizing certain attributes (e.g., scientist, engineer) to one sex or the other is taken as an index of implicit bias. Nosek, Banaji, and Greenwald (2002) found that people are generally quicker to associate men with science and women with literature, which is taken as an implicit bias against women in science, although they do note that their results may reflect, in part, the actual occupational sex differences in these areas. Even so, proponents argue that there could be a reciprocal relationship, whereby actual differences influence implicit biases that in turn dissuade girls and women from pursuing STEM fields (see Miller, Eagly, & Linn, 2015).

There is, however, vigorous debate regarding what exactly is being measured by these types of implicit tests (e.g., Greenwald, Nosek, Banaji, & Klauer, 2005; Greenwald, Banaji, & Nosek, 2015; Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013; Rothermund & Wentura, 2004) and whether they actually influence behavior (Blanton et al., 2009). Assuming the tests are actually measuring bias (e.g., sexism, racism), the relation between these implicit attitudes and actual behavior is small at best (e.g., Oswald et al., 2013), although proponents argue that these small effects add up over time (Greenwald et al., 2015). The ways in which implicit attitudes are thought to influence real-world outcomes include promoting stereotype threat (Miller et al., 2015) and microaggressions (Sue, 2010). As we noted above for stereotype threat, there are serious concerns about the ability to accurately measure microaggressions, whether they are related to implicit bias at all, if it is a valid concept, and whether 'victims' of microaggression suffer long-term consequences, among other concerns about the concept itself (see Lilienfeld, 2017). These issues have not stopped the development of yet another cottage industry for programs designed to make people aware of and to stop this 'aggression' on college campuses, in the workplace, and in daily life; an internet search for 'microaggression intervention' will provide many examples.

As with stereotype threat, the concepts of implicit bias and microaggression have gained such traction because they fit the narrative that inequalities of any kind are the result of some form of oppression; the entire narrative itself is a derivative of the postmodern spin on Marxism (Hicks, 2004). In many cases, explicit oppression is hard to find and thus the retort to unconscious bias and fleeting behaviors (microaggression) that continually 'assault' and undermine the 'victims'. In this case, the victims are girls' and women's aspirations toward and performance in STEM fields, especially engineering, computer science, and the physical sciences. The logical response to this narrative is the development of interventions to reduce stereotype threat, implicit bias, and microaggressions. But, what if these factors have much smaller effects on girls and women than proponents argue? The associated time and resources devoted to addressing these problems will have little or no long-term effect on girls' interest in or women's participation in inorganic STEM fields.

So, what is really going on? As with any life outcome that is complicated and unfolds over many years or decades, multiple factors likely contribute to the sex differences in interest and participation in STEM fields. Whatever the mix, proponents of stereotype threat, implicit bias, microaggression and related concepts expect that as societies become more equal, these forms of 'oppression' will diminish and boys and girls and men and women will become equal for most if not all non-physical traits, including participation in STEM (Hyde, 2005). Contrary to this hypothesis, we have recently found that countries renowned for gender equality show some of the largest sex differences in interest in and pursuit of STEM degrees (Stoet & Geary, 2018). For instance, Finland excels in gender equality (World Economic Forum, 2015), its adolescent girls outperform boys in science literacy, and it ranks near the top in European educational performance (Programme for International Student Assessment, 2016; https://nces.ed.gov/surveys/pisa/). With these high levels of educational performance and overall gender equality, Finland is poised to close the sex differences gap in STEM. Yet, Finland has one of the world's largest sex differences in college degrees in STEM fields, and Norway and Sweden, also leading in gender equality rankings, are not far behind. This is only the tip of the iceberg, as this general pattern of increasing sex differences with national increases in gender equality is found throughout the world, and not just for participation in STEM fields (e.g., Lippa, Collaer, & Peters, 2010).

The recent uptick in interest in concepts such as stereotype threat, implicit bias, and microaggression may be a reaction to this general phenomenon. If sex differences are the result of structural barriers (e.g., lack of employment opportunities), explicit sexism, and restricted educational opportunities, as they once were in many developed nations, then as these impediments fade into history, the sex differences attributed to them should fade as well. And, in fact some of them have faded and even reversed, such that more women than men attend and graduate from college and women now have structural advantages (e.g., hiring practices) in STEM fields (Ceci & Williams, 2015; Williams & Ceci, 2015). Despite these changes, many sex differences remain or have become larger over time. The latter are serious problems for anyone with strong beliefs about purely or largely social influences on sex differences and if the obvious social causes have been addressed, then there must be other, subtle oppressive factors that are causing these differences; enter stereotype threat, implicit bias, microaggression, and related concepts.

In any event, we propose that what is actually happening is that with economic development and advances in human rights, including gender equality, people are better able to pursue their individual interests and in doing so more basic sex differences are more fully expressed (Geary, 2010). With respect to STEM, these differences are related in part to student's interests and relative academic strengths. Sex differences in occupational interests, for instance, are large and well-documented, and reflect a more basic sex difference in interest in things versus people (Su, Rounds, & Armstrong, 2009). Men prefer occupations that involve working with things (e.g., engineering, mechanics) and abstract ideas (e.g., scientific theory) and women prefer working with and directly contributing to the wellbeing of others (e.g., physician, teacher). The sex difference in interest in people actually reflects a

more general interest in living things, which would explain why women who are interested in science are much more likely to pursue a career in biology or veterinary medicine than computer science (Lofstedt, 2003).

Although women and men are similar in intelligence, there are more specific cognitive and academic sex differences that influence educational and occupational choices (e.g., Geary, 1996). One of these differences is relative strengths in reading, mathematics, and science (Stoet & Geary, 2015). Students who are relatively better in reading-related areas (e.g., literature) than they are in science or mathematics (or visuospatial abilities), *independent* of their absolute level of performance relative to other students, are more likely to pursue college degrees in the humanities and enter non-science occupations, with the reverse for students who are relatively better in science and mathematics than literature (Humphreys, Lubinski, & Yao, 1993). This is where the results from Finland and elsewhere make sense. Although adolescent girls in Finland perform as well or better than their male peers in science, the gap is even larger in reading such that more Finnish girls have larger relative advantages in reading than science. Most adolescent boys in contrast are relatively better at science or mathematics than reading, independent of their absolute level of performance. Individuals with this pattern are likely to enter STEM areas, whether as research scientists or technicians, and there are more boys than girls with this pattern, worldwide (Stoet & Geary, 2015).

At the same time, there are substantive numbers of girls with relatively higher science or mathematics than reading achievement – 24% of Finnish girls – but proportionately fewer of these girls pursue STEM degrees than their male peers (Stoet & Geary, 2018). The gap between the number of adolescent girls with a STEM-biased academic pattern and the number of women who obtain a STEM degree in college is not likely due to stereotype threat, implicit bias, or related factors, because this gap increases with increases in national levels of gender equality. Early studies have shown that mathematically gifted women enter STEM fields less often than mathematically gifted men, not because of bias or microaggression, but because they have broader educational interests and thus consider a wider range of occupations than these men (Lubinski & Benbow, 1992). It seems to us that interventions focused on this group of girls (e.g., individual mentoring) holds much more promise for increasing the number of women in inorganic STEM professions than do currently vogue interventions that focus on rending the wider society of stereotypes, implicit bias, and microaggressions.

## References

Basford, T. E., Offermann, L. R., & Behrend, T. S. (2014). Do you see what I see? Perceptions of gender microaggressions in the workplace. *Psychology of Women Quarterly, 38*, 340–349.

Blanton, H., Jaccard, J., Klick, J., Mellers, B., Mitchell, G., & Tetlock, P. E. (2009). Strong claims and weak evidence: Reassessing the predictive validity of the IAT. *Journal of Applied Psychology, 94*, 567.

Ceci, S. J., & Williams, W. M. (2015). Women have substantial advantage in STEM faculty hiring, except when competing against more-accomplished men. *Frontiers in Psychology, 6*, e1532.

Del Río, C., & Alonso-Villar, O. (2015). The evolution of occupational segregation in the United States, 1940–2010: Gains and losses of gender–race/ethnicity groups. *Demography, 52*(3), 967–988.

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53*, 25–44.

Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effects on girls' mathematics performance. *Developmental Psychology, 49*, 1886–1897.

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences, 19*, 229–284.

Geary, D. C. (2010). *Male, female: The evolution of human sex differences* (2nd ed.). Washington, D.C.: American Psychological Association.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the implicit association test can have societally large effects. *Journal of Personality and Social Psychology, 108*, 553–561.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464.

Greenwald, A. G., Nosek, B. A., Banaji, M. R., & Klauer, K. C. (2005). Validity of the salience asymmetry interpretation of the implicit association test: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology: General, 134*, 420–425.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology, 97*, 17–41.

Hartocollis, A. (2016, June 26). Tutors see stereotypes and gender Bias in SAT. *Testers See None of the Above.* http://www.nytimes.com/2016/06/27/us/tutors-see-stereotypes-and-gender-bias-in-sat-testers-see-none-of-the-above.html?_r=1

Hicks, S. R. (2004). *Explaining postmodernism: Skepticism and socialism from Rousseau to Foucault*. Scholary Publishing, Inc. Tempe, AZ.

Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. Washington, D.C.: American Association of University Women.

Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist. *Journal of Applied Psychology, 78*, 250–261.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581–592.

Lilienfeld, S. O. (2017). Microaggressions: Strong claims, inadequate evidence. *Perspectives on Psychological Science, 12*, 138–169.

Lippa, R. A., Collaer, M. L., & Peters, M. (2010). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Archives of Sexual Behavior, 39*, 990–997.

Lofstedt, J. (2003). Gender and veterinary medicine. *The Canadian Veterinary Journal, 44*, 533–535.

Lubinski, D., & Benbow, C. P. (1992). Gender differences in abilities and preferences among the gifted: Implications for the math/science pipeline. *Current Directions in Psychological Science, 1*, 61–66.

Miller, D. I., Eagly, A. H., & Linn, M. C. (2015). Women's representation in science predicts national gender-science stereotypes: Evidence from 66 nations. *Journal of Educational Psychology, 107*, 631–644.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*, 101–115.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology, 105*, 171–192.

Picho, K., Rodriguez, A., & Finnie, L. (2013). Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *The Journal of Social Psychology, 153*, 299–333.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the implicit association test: Dissociating salience from associations. *Journal of Experimental Psychology: General, 133*, 139–165.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4–28.

Stoet, G., & Geary, D. C. (2012). Can stereotype threat explain the sex gap in mathematics performance and achievement? *Review of General Psychology, 16*, 93–102.

Stoet, G., & Geary, D. C. (2015). Sex differences in academic achievement are not related to political, economic, or social equality. *Intelligence, 48*, 137–151.

Stoet, G., & Geary, D. C. (2018). The gender equality paradox in STEM education. *Psychological Science, 29*, 581–593.

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people. *Psychological Bulletin, 135*, 859–884.

Sue, D. W. (2010). *Microaggressions in everyday life: Race, gender, and sexual orientation*. Hoboken, NJ: Wiley.

Thiederman, S. (2015). *3 keys to defeating unconscious bias*. San Diego, CA: Cross-Cultural Communications.

Walton, G. M., Logel, C., Peach, J. M., Spencer, S. J., & Zanna, M. P. (2015). Two brief interventions to mitigate a "chilly climate" transform women's experience, relationships, and achievement in engineering. *Journal of Educational Psychology, 107*, 468–485.

Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveal 2: 1 faculty preference for women on STEM tenure track. *Proceedings of the National Academy of Sciences USA, 112*, 5360–5365.

World Economic Forum. (2015). *The global gender gap report 2015*. Geneva, Switzerland: World Economic Forum.

# Chapter 16
# Groupthink in Sex and Pornography "Addiction": Sex-Negativity, Theoretical Impotence, and Political Manipulation


**Nicole Prause and D J Williams**

Concepts of excessive sexuality have existed for hundreds of years, but have recently turned profitable. The concept of profiting from treating sex as addictive was invented in the early 1980s with the publication of Carnes' (1983) clinical observations titled, *Out of the Shadows: Understanding Sexual Addiction.* Despite the fact that no science existed to support the model at the time, "addiction" was the first chosen framework. Speculatively, this model of sexual behavior would be most profitable: "addiction" treatments can command inpatient resources in contrast to typical time-limited, outpatient approaches for problems of compulsivity or relationship discord. Gradually, a diverse variety of academics, professionals, policymakers, and lay people have become increasingly concerned about sexual behavior that is commonly interpreted to be "out of control." While sexual "addiction" emerged largely due to cultural anxieties following the sexual revolution (Irvine, 1995), it gained momentum in large part due to its medicalization. Media accounts of celebrities who claimed to succumb to this supposed disorder fanned fashionable flames (Reay, Attwood, & Gooder, 2013). As the sex addiction industry became more firmly established, the target then widened to include viewing pornography, or more precisely, visual sexual stimuli (VSS). Currently, sex (and pornography)"addiction" are commonly discussed as separate, yet somewhat overlapping, clinical and political issues. However, the argument for the application of an addiction model to both sexual frequency and VSS rests on the same basic assumptions, shares the same logic, and is often promoted by the same believers.

The scientific method is designed to produce knowledge that is objective and valid. Science requires falsifiable hypotheses generated by the proposed model,

N. Prause
Liberos LLC, Los Angeles, CA, USA

D J Williams (✉)
Social Work, & Criminology, Idaho State University, Pocatello, ID, USA

Center for Positive Sexuality (Los Angeles), Los Angeles, CA, USA
e-mail: willdj@isu.edu



D. M. Allen, J. W. Howell (eds.), *Groupthink in Science*,
https://doi.org/10.1007/978-3-030-36822-7_16

then conducting carefully controlled studies with the ability to disprove (falsify) each hypothesis. This is a very high standard because every model prediction must hold true for the model to retain support. In the social and behavioral sciences, in particular, researchers must carefully consider broader social and cultural contexts that may influence the research process and interpretation of findings. Considering sociocultural influence on research questions requires careful exploration of potential extraneous variables, and alternative theories that could help explain patterns of behaviors. After such thorough examination, when data consistently fail to disprove the hypothesis, then support for the model is warranted. A model is never considered entirely "proven" because it is always subject to future falsification. In short, rigorous science, along with high quality scholarship more generally, demands considerable skepticism and critical analysis. Thus, the role of the scientist is primarily as a debunker, attempting to identify empirical fail points of proposed models. When one model fails, another, better-fitting model must be considered.

Consistent with other reviews (i.e., Ley, 2012, 2018; Ley, Prause, & Finn, 2014; Prause & Fong, 2015; Williams et al., 2017; Voros, 2009), we find that the addiction model as applied to sex and VSS viewing fails to meet scientific criteria for model support. Although only one hypothesis generated by the addiction model would need to be falsified to reject the model, many hypotheses generated by the addiction model have been falsified. These falsifications have been replicated by independent laboratories. Thus, "sex addiction" should not be considered a valid model, much less a diagnosis of pathology. A frequent objection by therapists is that debating the "addiction" model is merely becoming distracted by labels (see below). This reflects a basic misunderstanding of science. What you "call it" actually refers to the model being tested and defines how best to help.

Sexual scientists recognize the high complexity of frequent sexual behaviors and have parsed many models that could describe these behaviors, including non-pathology models (Walton, Cantor, Bhullar, & Lykins, 2017). In this chapter we focus on the scientific rejection of the "addiction" model of frequent sex and VSS viewing as contrasted by its perseverance in popular parlance. We hope to contribute to the current discussion concerning fundamental philosophical and methodological problems associated with the application of an addiction model, particularly as promoted by groupthink that runs afoul of basic principles of science. We draw attention to the impact of widespread sociohistorical sex-negativity, the need to consider broader theoretical explanations, and the political strategy for somewhat disparate institutions to adopt the veneer of science to promote their respective self-interests.

## Sex Negativity and Sociohistorical Considerations

Despite the fact that sexual norms, and thus also laws and moral judgments concerning sexuality, vary tremendously across cultures and historical time periods (Bullough, 1976; Hayes & Carpenter, 2012; Popovic, 2006), there has been a lack

of recognition of sociosexual diversity within sexology (Bhugra, Popelyuk, & McMullen, 2010). Bullough (1976) classified various cultures as more or less sex negative or sex positive. Sex-negativity is characterized by sexual asceticism, a narrow range of socially accepted sexual behaviors, lack of openness to sexuality, and sociosexual scripts preoccupied with risk and danger. Sex-positivity, or positive sexuality, acknowledges risk and danger, yet also recognizes the importance of sexual pleasure and wellbeing, embraces sexual diversity, and encourages open communication. Positive sexuality acknowledges personal and cultural diversity regarding sexuality and focuses less on sexual "deviance," and more on the ethics of various sexual practices (Williams, Christensen, & Capous-Desyllas, 2016).

There is little doubt that much of Western society, historically, has been thoroughly sex-negative (for an example, see Le Bodic, 2009 and Malan & Bullough, 2005 for a history of masturbation). American culture, in particular, continues to struggle with all types of sociosexual matters. The United States has been painfully slow to acknowledge and support the rights of lesbian, gay, bisexual, and transgender (LGBT) persons (i.e., Adam, 2003; Huebner, Rebchook, & Kegeles, 2004; Scott, 1998); support women's reproductive choices and provide access to contraception (i.e., Deckman & McTague, 2014; Harrison, 2005); and fully accept a range of consensual erotic and sexual practices (i.e., Ortmann & Sprott, 2012; Rubin, 1984). Furthermore, a review of contemporary U.S. sexual offending policy found that policy is terribly costly, sometimes increases injustice, and is largely fueled by myths rooted in widespread sex-negativity, rather than the large body of existing research (Williams, Thomas, & Prior, 2015).

It is not surprising that sexual literacy is a widespread problem. Sex education is mandated in schools in fewer than half of the U.S. states, and only 13 states require information to be medically accurate (Guttmacher Institute, 2012). At the same time, the federal government in recent decades has largely funded abstinence-only programs, particularly during the Bush administration, despite meta-analytic research showing that such programs are ineffective (Kirby, 2007). Sex education scholars have also pointed out that current sex education programs may unknowingly perpetuate a hegemonic sexuality with racial, class, and gender inequalities built into them (Connell & Elliott, 2009; Hobaica & Kwon, 2018; Hoefer & Hoefer, 2017). In focus groups concerning the effects of VSS viewing, a primary concern is that groupthink drives discussants to attempt to prove their righteousness by being critical of VSS (Iantaffi, Wilkerson, Grey, & Rosser, 2015).

Scientists and clinicians, of course, function within, and are influenced by, the broader sociohistorical context. In a climate of widespread sex-negativity, federal funding for scientific research on sexuality has generally been quite scarce, with virtually no funding for projects that consider positive possibilities of sexuality. Projects concerning sexuality at all that receive federal funding from the National Institutes of health have been uniquely attacked politically merely for possessing content on sexuality (Epstein, 2006). Curiously, it has only been recently that public health scholars have begun to consider seriously the potential psychosocial health benefits of sexuality and the importance of sexual pleasure (Anderson, 2013; Diamond & Huebner, 2012; Satcher, Hook III, & Coleman, 2015). When considering

health generally, scientists and clinicians have, for quite some time, followed the World Health Organization (WHO) recommendation that good health is more than simply the absence of disease, but includes positive dimensions, such as quality of life, life satisfaction, and overall wellbeing. However, the acceptance of similar positive constructs into definitions of sexual health has been slow. Moreover, while defining sexual health is shaped by sociohistorical events (Edwards & Coleman, 2004), there is often a lag time between accepted operational definitions of constructs and the widespread application of constructs within clinical practice.

## Here We Go Again! Changing Discourse from "Badness" to "Sickness"

Two decades ago, Irvine (1995) traced how the social process of medicalization led to the invention of sex addiction. This same social process has occurred previously with other notable sexual "disorders," such as masturbation and homosexuality. At the heart of medicalization is the use of language. While therapists argue that diagnostic labels assist validating patients' experiences, data show this is far from a universal experience, with as many diagnosed feeling devalued as helped by their label (Perkins et al., 2018). In their classic work on the medicalization of deviance, generally, Conrad and Schneider (1992) documented how discourses on deviant behavior have shifted from interpretations of "badness" to reinterpretation as "sickness." In considering a range of scholarship on sexuality, Hammack, Mayers, and Windell (2013) reported that sickness script changed in the 1970s to a "species" script following the removal of homosexuality in the DSM in 1973, and then to a "subject" script in the 1990s when scholarship diversified (including the emergence of queer theory). In their review on the interpretation of sexual deviance, De Block and Adriaens (2013) discuss the historical difficulties that the field of psychiatry has had, and continues to have, in classifying and understanding sexual behaviors. In considering sociohistorical issues and the diversification of scholarship, this has become more challenging. Indeed, specific terms do make a difference because of the scripts in which they are embedded. In addressing sexual variation, is there a different connotation between "deviance" and "diversity"?

## Helping Professions and Culturally Biased "Evidence"

The public may assume that contemporary helping professions, including psychology, counseling, social work, and marriage and family therapy, use interventions that are informed by a sound body of research and evidence. The American Counseling Association (ACA) *Code of Ethics* (2014, p. 10) states: "When providing services, counselors use techniques/procedures/modalities that are grounded in

theory and/or have an empirical or scientific foundation." Note, having a "scientific foundation" is an option, not a requirement, for counselor practice. Related fields try to avoid regulation with even weaker requirements. The American Psychological Association (APA) remains neutral, offering empirically supported treatment as one option, rather than a requirement (Elmore, 2016). A push by some psychologists to science-based interventions caused so much tension within APA that a schism formed and clinical science emerged (McFall, 1991). While valid debates exist concerning how to best implement ESTs, such as avoiding trademarked therapies (Rosen & Davison, 2003) and treatments less effective for minority clients (Bernal & Scharro-del-Rio, 2001), the case for distrusting clinical judgment over data remains extensive (Meehl, 1957; Miller, Spengler, & Spengler, 2015). Therapists' confidence in their own outcomes with patients typically far exceed their actual positive impact (Waller & Turner, 2016) and often do no better, or have outcomes even worse, than untrained paraprofessionals (Berman & Norton, 1985). Therapists raise many objections to following science-based treatments, including beliefs that feelings cannot be measured, beliefs that they are more important than the therapy used, and preferring to use their "gut" instead of evidence (Gyani, Shafran, Rose, & Lee, 2015).

Marriage and family therapy (MFT) practitioners especially rejected empirically supported interventions, with many refusing to leave their "clinical intuition" for science-backed treatments. A review of their flagship *Journal of Marital and Family Therapy* showed quantitative content, especially clinical trials, actually decreased from 2005 to 2014 (Parker, Chang, & Thomas, 2016). Specifically, MFT authors instructed researchers "should avoid attitudes that can reflect the belief that they know better than clinical practitioners who have been working in the field for decades" (Dattilio, Piercy, & Davis, 2013, p. 10). Unfortunately, longitudinal data show that years of experience as a therapist actually are associated with decreased efficacy with patients (Dunkle & Friedlander, 1996; Erekson, Janis, Bailey, Cattani, & Pedersen, 2017; Goldberg et al., 2016). MFTs continue to be disconnected from, and resistant to, implementing science-based treatments (Withers, Reynolds, Reed, & Holtrop, 2017). Indeed, science is not mentioned in any part of the MFT Commission on Accreditation (Crane, Wampler, Sprenkle, Sandberg, & Hovestadt, 2002). This is partially a self-selection problem, where MFT students select their program in large part due to a perceived fit with their personal religious beliefs (Hertlein & Lambert-Shute, 2007). However, the lack of training in human sexuality at such programs also appears to increase the problem.

The helping professions, as a whole, require very little, if any, training on human sexuality. For example, while the Council on Social Work Education (CSWE, 2008, 2015) includes sexual orientation (along with age, class, color, culture, gender, gender identity and expression, immigration status, political ideology, race, religion, and sex) in its statement on human diversity, there is no requirement for training on sexuality at any level (bachelor, master, doctoral) of education. A content analysis of popular social work textbooks found a glaring absence of discussion about sexual diversity (Prior, Williams, Zavala, & Milford, 2016). Further, most MFT faculty do not have any focused training in human sexuality (Zamboni & Zaid, 2017). As a

result, MFT comfort with sexual topics has not improved over decades (Dermer & Bachenberg, 2015). Specifically, a majority of MFT practitioners surveyed reported discomfort counseling homosexual clients. Perhaps not surprising, then, it is easy to see why many well-intentioned professionals, functioning in a longstanding socio-historical climate of sex-negativity, uncritically accept and promote an addiction framework of sexual behavior and VSS viewing, despite sociocultural biases, which helping professions supposedly oppose, inherent in sex/VSS addiction concepts and screening instruments (see Joannides, 2012; Williams, 2017). Of course, helping professionals are authority figures and are viewed as being experts on the issues for which they provide services. Unfortunately, this is not always true when it comes to matters pertaining to sexuality. When there are new opportunities to provide services (and profit), it can be easy for groupthink to occur and medicalization promoted by helping professionals to expand.

## Religion Masquerading as Public Health and Neuroscience

Religion is a significant force in the sex and VSS viewing addiction movement. Dominant Western religious organizations have a long history of opposition to various sexual practices (i.e., those that are not monogamous, married, vanilla; Rubin, 1984) and VSS viewing (Thomas, 2013). Recent research has found that there is a strong positive relationship between religiosity and perceived VSS addiction even when the actual amount of VSS viewing is controlled (Grubbs, Exline, Paragament, Hook, & Carlisle, 2015). In their review of the literature, Grubbs and Perry (2018) found that moral incongruence about VSS viewing is common and is associated with greater distress about VSS viewing, more frequently reported problems with VSS viewing, and an increased likelihood of perceived addiction to VSS viewing. Sociological studies by Thomas (2013, 2016) documented religious institutions' shifting narratives regarding the effects of VSS viewing from being a problem of social deviance (1950s and 1960s) to a problem of temptation and sin (1970s), and finally, now almost exclusively (beginning in the 1980s), to a problem of addiction that can have negative public health effects on society. Subsequently, using data from popular religious magazines combined with national survey data, Thomas, Alper, and Gleason (2017) have traced how religious anti-VSS viewing narratives apparently become internalized among those within such religious traditions to function as a form of self-fulfilling prophecy with respect to marital satisfaction. Some have noted that this has made some strange coalitions, such as anti-pornography feminists lecturing in religious spaces and filing anti-pornography legislation together (Whittier, 2014). Limited coalitions between traditionally oppositional groups serve to decrease issue-specific opposition (Pullum, 2017). In this case, therapists want to make money, religious groups want to regulate sexual expression, and feminists want to limit (perceived) harm to women. The movement regularly claims secular roots to the public, but these religious alliances have been

revealed repeatedly in both personal (e.g., Allen, 2015) and political (e.g., Campbell, 2018) biases.

The recent shifting of discourse from religious to public health (scientific) discourse is purposeful to persuade both public and professional opinion to accept the sex/VSS viewing model. The strategy is to make the addiction model appear to be constructed based on objective, scientific evidence. This, of course, reflects the same classic pattern of medicalization (Conrad & Schneider, 1992) and lends an "objective" measure of social control to unsanctioned sexual behavior (Voros, 2009).

## *Sexy Neuroscience*

Attempts to appeal to authority when health is the topic make scientists the authority of claimed knowledge. Being told that scientists completely understand a phenomenon has been shown to increase the layperson's confidence in their own (inaccurate) knowledge (Sloman & Rabb, 2016). The field of neuroscience is especially widely touted for its documented ability to deceive consumers of health information. Viewing brain images in the context of health statements increases untrained individuals' beliefs in the information presented (McCabe & Castel, 2008), and this occurs without increasing their actual knowledge of neuroscience (Ikeda, Kitagami, Takahashi, Hattori, & Ito, 2013). Others have suggested that it is not the brain images per se that increase false confidence, but rather the presence of any neuroscience-sounding information, whether or not it was relevant to the research described (Hook & Farah, 2013). Thus, confidence may be most likely bolstered by the mention of neuroscience concepts when the actual science is most weak. This is a lucrative strategy. Using brain information to push addiction models has been shown to increase acceptance of treatment (see Figure 1 in Racine, Sattler, & Escande, 2017). Sex addiction clinicians appear anxious to appeal to neuroscience authority. The International Institute for the Treatment of Trauma and Addictions, an organization that licenses sex addiction therapists, advertised a talk on the "neuroscience" of sex by a speaker who actually was not a neuroscientist (IITAPllc, 2014). Rather, the speaker had self-published his only text on the topic for the Latter Day Saints' concerning how to use religion to overcome the evils of pornography. He later published a letter to the editor claiming to critique our study, which was so bizarre, rambling, and obviously uninformed about basic principles of neuroscience that we declined the journal's offer to respond to it. Climate scientists have faced similar challenges from the presentation of fake experts (Hansson, 2018).

There are conditions under which this bias may be reduced. Studies in which the participant was encouraged to question the presented neuroscience, such as using descriptions like "Can Brain Scans Detect Criminals?" reduced the bias to accept information presented with brain images (Schweitzer, Baker, & Risko, 2013). However, participants were less likely to believe direct critiques of neuroscience data rather than glowing, positive reviews of neuroscience data (Popescu, Thompson, Gayton, & Markowski, 2016). Where scientists accurately characterize data as

having falsified the addiction model of sex, activists confidently claim that the addiction model is "proven," despite that science can only support a model. Combine poor public discrimination of neuroscience evidence with confirmation bias of a sex-negative society, and it is not difficult to understand the political traction of various groups that are promoting such a false narrative.

The primacy of brain data is a problem that extends into the field of neuroscience. Descriptions of neuroscience data as "underlying" or "explaining" sexual behaviors represents a classic error of biological reductionism. In reality, science is integrative, with biology, behavior, social, and other levels of analysis often equally important in model testing (Cacioppo, 2002). The best model holds up across these levels of analysis. This is partly why psychology has been anxious to grab the designation of the "hub science" that can best integrate these sources of information (Cacioppo, 2007). The ability to document differences in proposed groups by brain activations provides no evidence that a particular group necessarily has a disease.

## Addiction Is the Wrong Model

While some people clearly are distressed by their sexual behaviors, it is important to identify the best model. The best model is one that best characterizes and predicts future behaviors. Thus, there are many models of high-frequency sexual behaviors. These include a number of non-pathological models (Walton et al., 2017), including the high sex drive and/or social shame model. These are empirically separable (Prause, 2017). While falsification of behavior models is a core tenet of science, it bears explanation. The therapists claiming to treat "sex addicts" describe differentiating models as irrelevant for treatment and reflecting merely different "names" for the same behaviors (Carnes & Love, 2017). Such fundamental misunderstandings of science are of concern for the type of care patients are likely to receive. Indeed, there is currently no random-assignment, controlled trial for sex or porn addiction as of this writing. Websites concerning "porn addiction" are especially likely, relative to other behavioral issue websites, to recommend religious absolution and complete abstinence as a goal (Rodda, Booth, Vacaru, Knaebe, & Hodgins, 2018). The most popular conceptualization by clinicians has been the "sex addiction" model, which is curious given that it has the weakest empirical support.

The specifics of an addiction model can, of course, vary a bit between scientists. However, most scientists agree that key features of any addiction include compulsions to seek the drug/behavior, a loss of control of the behavior or consumption, a withdrawal state (Koob & Le Moal, 2008), involvement of neural reward systems, and neuroadaptations over time that promote craving over liking (Robinson & Berridge, 2000). While an addiction model includes components of compulsivity and impulsivity, those ("compulsion" and "impulsivity") also are recognized as separable, distinct models from addiction (Prause, 2017).

By applying the falsification criterion to models of frequent sexual behaviors, the "addiction" model has been falsified (Prause, Steele, Staley, Sabatinelli, & Hajcak,

2016; Prause, Janssen, Georgiadis, Finn, & Pfaus, 2017). That is, several of the predictions made by an addiction model have failed in experiments. These experiments have been replicated and extended by independent laboratories, which is the gold standard for falsification.

Both the American Psychiatric Association (2013) and the World Health Organization International (WHO) specifically excluded "sex addiction" from their nomenclature (within *Diagnostic and Statistical Manual of Mental Disorders*, or DSM, and the *International Classification of Diseases*, or ICD). "Porn addiction" also was excluded from the ICD-11 (Grant et al., 2014). Notably, the ICD-11 is considering whether or not to add "compulsive sexual behavior" at this time. ICD currently requires ruling out "Distress that is entirely related to moral judgments and disapproval about sexual impulses." As no study to date has ever tested any patient sample that meets these requirements, it is unclear to whom the diagnosis would refer. From the first nationally representative study, 2.3% of men and 0.2% of women in the Netherlands reported feeling that they might be sexually compulsive (National Institute for Public Health, 2017). Given that this assessment did not rule out individuals with concerns due to moral judgments, it appears likely that such problems may not be experienced by any portion of the population. More succinctly, such tiny numbers appear within the error variance of self-description. Notably, the oft-repeated prevalence guess of one sex and pornography addiction therapist for these difficulties (Carnes, 2013) turned out to be 2.6 (men) to 30 (women) times higher than suggested by actual data from nationally representative samples.

Perhaps the most common scientific misperception pushed by anti-pornography organizations is that dopamine involvement is the same as addiction (Ley, 2018). For example, alarming titles such as "Technology gives us dopamine...highly addictive!" (Sprout, 2017) and "Sex releases the highest levels of dopamine naturally available, equal to morphine & nicotine" (Wilson, 2018) are touted to gain political support for an addiction model. Both statements are false. Dopamine is involved in many functions, including learning, salience, and movement (Schultz, Stauffer, & Lak, 2017). Dopamine is not specific to addiction. Further, dopamine has never been compared by titers with substances; in fact, null-hypothesis statistics could never support the conclusion that conditions are "equal." Certainly, there is strong evidence that increases in dopamine availability increase sexual behaviors just as sexual behaviors themselves increase the activity of dopamine. These are necessary, but not sufficient, conditions for addiction (see above). Dopamine activity would need to be involved to support an addiction model, but dopamine is altered in many behaviors with no relationship to any proposed addiction.

Withdrawal hypotheses appear to lack empirical support. Even with substances, withdrawal is not consistently a required feature, such as for inhalants (Hasin et al., 2013). Similarly, behavioral addiction clinicians sometimes advocate removing the requirement of withdrawal for behaviors (Van Rooij & Prause, 2014). However, clinicians have argued that "sex addiction" patients exhibit withdrawal. For example, Goodman (2001) argued that withdrawal is a component of "sex addiction" but that withdrawal need not be evidenced physiologically. In direct contradiction, the withdrawal symptoms reported by other clinicians (Karila et al., 2014) include only

and explicitly physiological symptoms, with 70% of patients claiming experiences of "nervousness, insomnia, sweating, nausea, increased heart rate, shortness of breath, and fatigue." The field of psychophysiology is well-equipped to document all of these claimed symptoms; yet none have been documented to date. Given that there also are currently no data on human sexual deprivation states in non-pathology, this hypothesis from the addiction model can reasonably be described as having no supportive data.

## Conclusion

Concerns around sexual behavior, including sex film viewing, appear largely driven by social forces. These forces include monetary gain (i.e., therapists and politicians), religious (i.e., Latter Day Saints and evangelicals), and ideological (i.e., feminists). To reach the goal of pathologizing these sexual behaviors, such groups have conspired to appropriate a false framework of "health" behaviors, which requires promoting an appearance of science. We have demonstrated that such collaborative adversarial movements (Whittier, 2014) led to gross overestimates of prevalence, basic misunderstanding of scientific model testing, mischaracterizations of neuroscience, appeals to fake authorities, and intentional disregard of disconfirming data. In fact, data suggest the best thing for individuals who report distress about their sexual behaviors is likely to do nothing. Curiously, a study of individuals who believed that they were "sex addicts" found that 100% of women ($N = 68$) and 95% of men ($N = 167$) spontaneously resolved their concerns without treatment over a 5-year period, and most were resolved within the first year of expressing the concern (Konkolÿ, Thege, Woodin, Hodgins, & Williams, 2015).

So how do we respect some individuals' distress about their own particular sexual behaviors given the current socio-cultural situation? First, we use standardized, validated assessments of accepted diagnoses. For example, depression is mistakenly described as "comorbid" with "porn addiction" where a primary diagnosis of depression is likely more appropriate and parsimonious. Many empirically supported depression treatments exist that accommodate sexual features, but no sex addiction ESTs exist. Second, psychoeducation is essential. Education is an important component of most sexual interventions. However, due to widespread sex negativity and poor sex education in the United States, there is extensive misinformation on the Internet, especially regarding what is "normal." Third, advocate for patients who are being misled, such as by calling attention to clinicians who refuse to base treatments on rigorous science (McFall, 1991). Patients may struggle to distinguish between qualified clinicians and those who are simply reproducing sex-negative discourses of pathology via neuroscience jargon. Also, many patients appear unaware that clinicians are not required to provide treatments with any scientific support. Scientists engaging in social media can provide information more directly to people with concerns (Bik & Goldstein, 2013). For those rightfully concerned about organized social attacks to providing this information online, it is useful to

consult guides that excuse scientists from corresponding with activists online (Lewandowsky & Bishop, 2016).

Unfortunately, groupthink on the topic of sex and pornography "addiction" is surprisingly common. There remains a glaring need for scientists and practitioners to remember that Western society remains saturated in a socio-cultural climate of sex-negativity. Proponents of the sex/pornography addiction movement are (often intentionally) influenced by their own broader interests (i.e., monetary, religious, ideological). Current social scripts concerning commonly disapproved sexual behaviors and identities reflect a long history of following a "badness" (religious) to "sickness" (public health) central theme. Finally, as we have documented herein, actual controlled, peer-reviewed neuroscientific investigations fail to support an addiction model.

## References


Adam, B. D. (2003). The defense of marriage act and American exceptionalism: The "gay marriage" panic in the United States. *Journal of the History of Sexuality, 12*, 259–276.

Allen, S. (2015, October 20). *"Porn kills love": Mormons' anti-smut crusade*. Retrieved April 19, 2018, from https://www.thedailybeast.com/articles/2015/10/20/porn-kills-love-mormons-anti-smut-crusade

American Counseling Association. (2014). *Code of ethics*. Alexandria, VA: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. Washington, D.C.: American Psychiatric Pub.

Anderson, R. M. (2013). Positive sexuality and its impact on overall well-being. *Bundesgesundheitsblatt, 56*, 208–214. https://doi.org/10.1007/s00103-012-1607-z

Berman, J. S., & Norton, N. C. (1985). Does professional training make a therapist more effective? *Psychological Bulletin, 98*, 401–407. https://doi.org/10.1037/0033-2909.98.2.401

Bernal, G., & Scharro-del-Rio, M. R. (2001). Are empirically supported treatments valid for ethnic minorities? Toward an alternative approach for treatment research. *Cultural Diversity & Ethnic Minority Psychology, 7*(4), 328–342.

Bhugra, D., Popelyuk, D., & McMullen, I. (2010). Paraphilias across cultures: Contexts and controversies. *Journal of Sex Research, 47*, 242–256. https://doi.org/10.1080/00224491003699833

Bik, H. M., & Goldstein, M. C. (2013). An introduction to social media for scientists. *PLoS Biology, 11*(4), e1001535.

Bullough, V. L. (1976). *Sexual variance in society and history*. Chicago, IL: University of Chicago Press.

Cacioppo, J. (2007). Psychology is a hub science. *APS Observer, 20*(8).

Cacioppo, J. T. (2002). Social neuroscience: Understanding the pieces fosters understanding the whole and vice versa. *The American Psychologist, 57*(11), 819–831.

Campbell, A. (2018, February 8). *Kansas senate uses fake science to declare porn a public health crisis*. Retrieved April 19, 2018, from https://www.huffingtonpost.com/entry/kansas-senate-porn-resolution_us_5a7c5bc9e4b0c6726e101d66

Carnes, P. (1983). *Out of the shadows: Understanding sexual addiction*. Minneapolis, MN: CompCare.

Carnes, P. (2013). *Don't call it love: Recovery from sexual addiction*. New York, NY: Bantam.

Carnes, S., & Love, T. (2017). Separating models obscures the scientific underpinnings of sex addiction as a disorder. *Archives of Sexual Behavior, 46*(8), 2253–2256.

Connell, C., & Elliott, S. (2009). Beyond the birds and the bees: Learning inequality through sexuality education. *American Journal of Sexuality Education, 4*, 83–102. https://doi.org/10.1080/15546120903001332

Conrad, P., & Schneider, J. W. (1992). *Deviance and medicalization: From badness to sickness*. Philadelphia, PA: Temple University Press.

Council on Social Work Education (CSWE). (2008). *Educational policy and educational standards*. Alexandria, VA: Author.

Council on Social Work Education (CSWE). (2015). *Final 2015 educational policy (EP)*. Alexandria, VA: Author.

Crane, D. R., Wampler, K. S., Sprenkle, D. H., Sandberg, J. G., & Hovestadt, A. J. (2002). The scientist-practitioner model in marriage and family therapy doctoral programs: Current status. *Journal of Marital and Family Therapy, 28*(1), 75–83.

Dattilio, F. M., Piercy, F. P., & Davis, S. D. (2013). The divide between "evidenced-based" approaches and practitioners of traditional theories of family therapy. *Journal of Marital and Family Therapy, 40*(1), 5–16.

De Block, A., & Adriaens, P. R. (2013). Pathologizing sexual deviance: A history. *Journal of Sex Research, 50*, 276–298. https://doi.org/10.1080/00224499.2012.738259

Deckman, M., & McTague, J. (2014). Did the "war on women" work? Women, men, and the birth control mandate in the 2012 presidential election. *American Politics Research, 43*, 3–26. https://doi.org/10.1177/1532673X14535240

Dermer, S., & Bachenberg, M. (2015). The importance of training marital, couple, and family therapists in sexual health. *Australian and New Zealand Journal of Family Therapy, 36*(4), 492–503.

Diamond, L. M., & Huebner, D. M. (2012). Is sex good for you? Rethinking sexuality and health. *Social and Personality Psychology Compass, 6*, 54–69. https://doi.org/10.1111/j.1751-9004.2011.00408.x

Dunkle, J. H., & Friedlander, M. L. (1996). Contribution of therapist experience and personal characteristics to the working alliance. *Journal of Counseling Psychology, 43*, 456.

Edwards, W. M., & Coleman, E. (2004). Defining sexual health: A descriptive overview. *Archives of Sexual Behavior, 33*, 189–195.

Elmore, A. (2016). Empirically supported treatments: Precept or percept? *Professional Psychology, Research and Practice, 47*(3), 198–205.

Epstein, S. (2006). The new attack on sexuality research: Morality and the politics of knowledge production. *Sexuality Research and Social Policy Journal of NSRC, 3*, 1–12. https://doi.org/10.1525/srsp.2006.3.1.01

Erekson, D. M., Janis, R., Bailey, R. J., Cattani, K., & Pedersen, T. R. (2017). A longitudinal investigation of the impact of psychotherapist training: Does training improve client outcomes? *Journal of Counseling Psychology, 64*(5), 514–524.

Goldberg, S. B., Rousmaniere, T., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., et al. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology, 63*(1), 1–11.

Goodman, A. (2001). What's in a name? Terminology for designating a syndrome of driven sexual behavior. *Sexual Addiction and Compulsivity, 8*(3–4), 191–213.

Grant, J. E., Atmaca, M., Fineberg, N. A., Fontenelle, L. F., Matsunaga, H., Reddy, Y. C. J., et al. (2014). Impulse control disorders and "behavioural addictions" in the ICD-11. *World Psychiatry: Official Journal of the World Psychiatric Association, 13*(2), 125–127.

Grubbs, J. B., Exline, J. J., Paragament, K. I., Hook, J. N., & Carlisle, R. D. (2015). Transgression as addiction: Religiosity and moral disapproval as predictors of perceived addiction to pornography. *Archives of Sexual Behavior, 44*, 125–136. https://doi.org/10.1007/s10508-013-0257-z

Grubbs, J. B., & Perry, S. L. (2018). Moral incongruence and pornography use: A critical review and integration. *Journal of Sex Research* (online first). https://doi.org/10.1080/00224499.2018.1427204

Guttmacher Institute. (2012). *State policies in brief: Sex and HIV education*. New York, NY: Author.

Gyani, A., Shafran, R., Rose, S., & Lee, M. J. (2015). A qualitative investigation of therapists' attitudes towards research: Horses for courses? *Behavioural and Cognitive Psychotherapy, 43*, 436–448.

Hammack, P. L., Mayers, L., & Windell, E. P. (2013). Narrative, psychology and the politics of sexual identity in the United States: From "sickness" to "species" to "subject". *Psychology and Sexuality, 4*, 219–243. https://doi.org/10.1080/19419899.2011.621131

Hansson, S. O. (2018). Dealing with climate science denialism: Experiences from confrontations with other forms of pseudoscience. *Climate Policy, 1*, 1–9.

Harrison, T. (2005). Availability of emergency contraception: A survey of hospital emergency department staff. *Annals of Emergency Medicine, 46*(2), 105–110. https://doi.org/10.1016/j.annemergmed.2005.01.017

Hasin, D. S., O'Brien, C. P., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., et al. (2013). DSM-5 criteria for substance use disorders: Recommendations and rationale. *The American Journal of Psychiatry, 170*(8), 834–851.

Hayes, S., & Carpenter, B. (2012). Out of time: The moral temporality of sex, crime, and taboo. *Critical Criminology, 20*, 141–152. https://doi.org/10.1007/s10612-011-9130-3

Hertlein, K. M., & Lambert-Shute, J. (2007). Factors influencing student selection of marriage and family therapy graduate programs. *Journal of Marital and Family Therapy, 33*(1), 18–34.

Hobaica, S., & Kwon, P. (2018). "This is how you hetero:" sexual minorities in heteronormative sex education. *American Journal of Sexuality Education, 12*, 423–450. https://doi.org/10.1080/15546128.2017.1399491

Hoefer, S. E., & Hoefer, R. (2017). Worth the wait? The consequences of abstinence-only sex education for marginalized students. *American Journal of Sexuality Education, 12*, 257–276. https://doi.org/10.1080/15546128.2017.1359802

Hook, C. J., & Farah, M. J. (2013). Look again: Effects of brain images and mind-brain dualism on lay evaluations of research. *Journal of Cognitive Neuroscience, 25*(9), 1397–1405.

Huebner, D. M., Rebchook, G. M., & Kegeles, S. M. (2004). Experiences of harassment, discrimination, and physical violence among young gay and bisexual men. *American Journal of Public Health, 94*, 1200–1203. https://doi.org/10.2105/AJPH.94.7.1200

Iantaffi, A., Wilkerson, J. M., Grey, J. A., & Rosser, B. R. (2015). Acceptability of sexually explicit images in HIV prevention messages targeting men who have sex with men. *Journal of Homosexuality, 62*, 1345–1358. https://doi.org/10.1080/00918369.2015.1060066

IITAPllc. (2014, February 21). *Dr. Don Hilton speaking about the neuroscience of pornography #addiction. #IITAPSymposium*. Retrieved from https://twitter.com/iitapllc/status/436897536075239424

Ikeda, K., Kitagami, S., Takahashi, T., Hattori, Y., & Ito, Y. (2013). Neuroscientific information bias in metacomprehension: The effect of brain images on metacomprehension judgment of neuroscience research. *Psychonomic Bulletin & Review, 20*(6), 1357–1363.

Irvine, J. M. (1995). Reinventing perversion: Sex addiction and cultural anxieties. *Journal of the History of Sexuality, 5*, 429–450.

Joannides, P. (2012). The challenging landscape of problematic sexual behaviors, including "sexual addiction" and "hypersexuality". In P. Kleinplatz (Ed.), *New directions in sex therapy: Innovations and alternatives* (pp. 69–84). New York, NY: Routledge.

Karila, L., Wéry, A., Weinstein, A., Cottencin, O., Petit, A., Reynaud, M., & Billieux, J. (2014). Sexual addiction or hypersexual disorder: different terms for the same problem? A review of the literature. *Current pharmaceutical design, 20*(25), 4012–4020.

Kirby, D. (2007). Abstinence, sex, and STD/HIV education programs for teens: Their impact on sexual behavior, pregnancy, and sexually transmitted disease. *Annual Review of Sex Research, XVIII*, 143–177.

KonkolÿThege, B., Woodin, E. M., Hodgins, D. C., & Williams, R. J. (2015). Natural course of behavioral addictions: A 5-year longitudinal study. *BMC Psychiatry, 15*, 4.

Koob, G. F., & Le Moal, M. (2008). Review. Neurobiological mechanisms for opponent motivational processes in addiction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 363*(1507), 3113–3123.

Le Bodic, C. (2009). Masturbation and therapy: The example of treatment for perpetrators of sexual abuse. *Sexologies, 18*, 255–258. https://doi.org/10.1016/j.sexol.2009.09.003

Lewandowsky, S., & Bishop, D. (2016). Research integrity: Don't let transparency damage science. *Nature, 529*(7587), 459–461.

Ley, D. (2012). *The myth of sex addiction*. Lanham, MD: Rowman & Littlefield.

Ley, D. (2018). The pseudoscience behind public health crisis legislation. *Porn Studies* (online first). https://doi.org/10.1080/23268743.2018.1435400

Ley, D., Prause, N., & Finn, P. (2014). The emperor has no clothes: A review of the "pornography addiction" model. *Current Sexual Health Reports, 6*, 94–105. https://doi.org/10.1007/s11930-014-0016-8

Malan, M. K., & Bullough, V. (2005). Historical development of new masturbation attitudes in Mormon culture: Silence, secular conformity, counterrevolution, and emerging reform. *Sexuality and Culture, 9*(4), 80–127.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition, 107*(1), 343–352.

McFall, R. M. (1991). Manifesto for a science of clinical psychology. *The Clinical Psychologist, 44*(6), 75–88.

Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology, 4*(4), 268–273.

Miller, D. J., Spengler, E. S., & Spengler, P. M. (2015). A meta-analysis of confidence and judgment accuracy in clinical decision making. *Journal of Counseling Psychology, 62*(4), 553–567.

National Institute for Public Health and the Environment. (2017). Groupthink in Sex and Pornography "Addiction": Sex-Negativity, Theoretical Impotence, and Political Manipulation. In H. de Graaf Ciel Wijsen, (Ed.), *Sexual health in the Netherlands*. New York: Springer.

Ortmann, D. M., & Sprott, R. (2012). *Sexual outsiders: Understanding BDSM sexualities and communities*. Lanham, MD: Rowman & Littlefield.

Parker, E. O., Chang, J., & Thomas, V. (2016). A content analysis of quantitative research in journal of marital and family therapy: A 10-year review. *Journal of Marital and Family Therapy, 42*(1), 3–18.

Perkins, A., Ridler, J., Browes, D., Peryer, G., Notley, C., & Hackmann, C. (2018). Experiencing mental health diagnosis: A systematic review of service user, clinician, and carer perspectives across clinical settings. *The Lancet Psychiatry*. https://doi.org/10.1016/s2215-0366(18)30095-6

Popescu, M., Thompson, R. B., Gayton, W., & Markowski, V. (2016). A reexamination of the neurorealism effect: The role of context. *Journal of Science Communication, 15*(6), A01.

Popovic, M. (2006). Psychosexual diversity as the best representation of human normality across cultures. *Sexual and Relationship Therapy, 21*, 171–186. https://doi.org/10.1080/14681990500358469

Prause, N. (2017). Evaluate models of high-frequency sexual behaviors already. *Archives of Sexual Behavior, 46*(8), 2269–2274.

Prause, N., & Fong, T. (2015). The science and politics of sex addiction research. In L. Comella & S. Tarrant (Eds.), *New views on pornography: Sexuality, politics, and the law*. Santa Barbara, CA: Praeger.

Prause, N., Janssen, E., Georgiadis, J., Finn, P., & Pfaus, J. (2017). Data do not support sex as addictive. *The Lancet Psychiatry, 4*(12), 899.

Prause, N., Steele, V. R., Staley, C., Sabatinelli, D., & Hajcak, G. (2016). Prause et al. (2015) the latest falsification of addiction predictions. *Biological Psychology, 120*, 159–161.

Prior, E. E., Williams, D. J., Zavala, T., & Milford, J. (2016). *What do(n't) American undergraduate social work students learn about sex? A content analysis of sex positivity*. University of Windsor, http://www1.uwindsor.ca/criticalsocialwork/HBSEtextbooks

Pullum, A. (2017). Foul Weather Friends: Enabling Movement Alliance through an Intentionally Limited Coalition. *Social Currents, 5*(3), 228–243. https://doi.org/10.1177/2329496517725329

Racine, E., Sattler, S., & Escande, A. (2017). Free will and the brain disease model of addiction: The not so seductive allure of neuroscience and its modest impact on the attribution of free will to people with an addiction. *Frontiers in Psychology, 8*, 1850.

Reay, B., Attwood, N., & Gooder, C. (2013). Inventing sex: The short history of sex addiction. *Sexuality and Culture, 17*, 1–19. https://doi.org/10.1007/s12119-012-9136-3

Robinson, T. E., & Berridge, K. C. (2000). The psychology and neurobiology of addiction: An incentive–sensitization view. *Addiction, 95*(8), 91–117.

Rodda, S. N., Booth, N., Vacaru, M., Knaebe, B., & Hodgins, D. C. (2018). Behaviour change strategies for internet, pornography and gaming addiction: A taxonomy and content analysis of professional and consumer websites. *Computers in Human Behavior, 84*, 467–476.

Rosen, G. M., & Davison, G. C. (2003). Psychology should list empirically supported principles of change (ESPs) and not credential trademarked therapies or other treatment packages. *Behavior Modification, 27*(3), 300–312.

Rubin, G. (1984). Thinking sex: Notes for a radical theory of the politics of sexuality. In C. Vance (Ed.), *Exploring female sexuality* (pp. 267–319). Boston, MA: Routledge.

Satcher, D., Hook III, E. W., & Coleman, E. (2015). Sexual health in America: Improving patient care and public health. *JAMA, 314*, 765–766.

Schultz, W., Stauffer, W. R., & Lak, A. (2017). The phasic dopamine signal maturing: From reward via behavioural activation to formal economic utility. *Current Opinion in Neurobiology, 43*, 139–148.

Schweitzer, N. J., Baker, D. A., & Risko, E. F. (2013). Fooled by the brain: Re-examining the influence of neuroimages. *Cognition, 129*(3), 501–511.

Scott, J. (1998). Changing attitudes toward sexual morality: A cross-national comparison. *Sociology, 32*, 815–845.

Sloman, S. A., & Rabb, N. (2016). Your understanding is my understanding: Evidence for a community of knowledge. *Psychological Science, 27*(11), 1451–1460.

StaciSprout. (2017, September 24). *Technology gives us dopamine...highly addictive*! Retrieved from https://twitter.com/Stacisprout/status/912173274599014400

Thomas, J. N. (2013). Outsourcing moral authority: The internal secularization of evangelicals' anti-pornography narratives. *Journal for the Scientific Study of Religion, 52*, 182–206. https://doi.org/10.1111/jssr.12052

Thomas, J. N. (2016). The development and deployment of the idea of pornography addiction within American evangelicalism. *Sexual Addiction and Compulsivity, 23*, 182–195. https://doi.org/10.1080/10720162.2016.1140603

Thomas, J. N., Alper, B. A., & Gleason, S. A. (2017). Anti-pornography narratives as self-fulfilling prophecies: Religious variation in the effect that pornography viewing has on the marital happiness of husbands. *Review of Religious Research, 59*, 471–497. https://doi.org/10.1007/s13644-107-0301-x

Van Rooij, A., & Prause, N. (2014). A critical review of "internet addiction" criteria with suggestions for the future. *Journal of Behavioral Addictions, 3*(4), 203–213.

Voros, F. (2009). The invention of addiction to pornography. *Sexologies, 18*, 243–246. https://doi.org/10.1016/j.sexol.2009.09.007

Waller, G., & Turner, H. (2016). Therapist drift redux: Why well-meaning clinicians fail to deliver evidence-based therapy, and how to get back on track. *Behaviour Research and Therapy, 77*, 129–137.

Walton, M. T., Cantor, J. M., Bhullar, N., & Lykins, A. D. (2017). Hypersexuality: A critical review and introduction for the "sexhavior cycle". *Archives of Sexual Behavior, 46*, 2231–2251. https://doi.org/10.1007/s10508-017-0091-8

Whittier, N. (2014). Rethinking coalitions: Anti-pornography feminists, conservatives, and relationships between collaborative adversarial movements. *Social Problems, 61*(2), 175–193.

Williams, D. J. (2017). The framing of frequent sexual behavior and/or pornography viewing as addiction: Some concerns for social work. *Journal of Social Work, 17*, 616–623. https://doi.org/10.1177/1468017316644701

Williams, D. J., Christensen, M. C., & Capous-Desyllas, M. (2016). Social work practice and sexuality: Applying a positive sexuality model to enhance diversity and resolve problems. *Families in Society, 97*, 287–294. https://doi.org/10.1606/1044-3894.2016.97.35

Williams, D. J., Thomas, J. N., & Prior, E. E. (2015). Moving full-speed ahead in the wrong direction? A critical examination of US sex offender policy from a positive sexuality model. *Critical Criminology, 23*, 277–294. https://doi.org/10.1007/s10612-015-9270-y

Williams, D. J., Thomas, J. N., Prior, E. E., Wright, S., Sprott, R., et al. (2017). Sex and pornography "addiction": An official position statement of the Center for Positive Sexuality (CPS), National Coalition for sexual freedom (NCSF), and the alternative sexualities Health Research Alliance (TASHRA). *Journal of Positive Sexuality, 3*, 40–43.

Wilson, G. (2018, March 9). *Sex releases the highest levels of dopamine naturally available, equal to morphine & nicotine. Sex & addictive drugs (meth, cocaine) activate the same reward center neurons, which differ from other natural rewards. He should read Hilton's article*. Retrieved from https://twitter.com/YourBrainOnPorn/status/972119573502701568

Withers, M. C., Reynolds, J. E., Reed, K., & Holtrop, K. (2017). Dissemination and implementation research in marriage and family therapy: An introduction and call to the field. *Journal of Marital and Family Therapy, 43*(2), 183–197.

Zamboni, B. D., & Zaid, S. J. (2017). Human sexuality education in marriage and family therapy graduate programs. *Journal of Marital and Family Therapy, 43*(4), 605–616.

# Chapter 17
# The Tyranny of the Normal Curve: How the "Bell Curve" Corrupts Educational Research and Practice


**Curt Dudley-Marling**

*Does the good of the many outweigh the good of the one? –*

(Spock's mother, Star Trek IV [Nimoy, n.d.]*)*

The idea that human behavior distributes more or less "normally" along the lines of a bell-shaped curve (the *normal curve*) has achieved the level of common sense in American popular culture as well as educational and social science research. It is generally assumed that various human traits cluster around the mean of a more or less *normal* distribution and, for many traits and abilities, people may be defined in terms of their relationship to the mean (or average). For traits like body size and temperament and mental health, for instance, average is typically presented as the ideal (i.e., normal) and people who fall outside the boundaries of normal for these traits are at risk for being stigmatized as *abnormal*. For traits like intelligence, appearance and athleticism, on the other hand, above average is most desirable while below average for these traits may lead to lower social status. Overall, the lens of normality affects how we see ourselves and others and how we organize our institutions including the institution of schooling.

The ideology of the normal curve is a foundational principle of modern schooling. The assumption that human behavior tends to fall along the boundaries of a bell curve, with most people clustering around the mean, affects how schools are organized, how students are taught and evaluated, who is included (and excluded) from the "normal" classroom curriculum, and how educational research is conducted and interpreted—particularly how educational research is used to inform classroom practice. There are, however, fundamental problems relying on norm-based research as a basis for educational decision-making. For starters, only truly random events distribute *normally*, and the behavior of human beings, unlike the roll of the dice or the flip of a coin, is never truly random. Moreover, making claims about individuals

C. Dudley-Marling (✉)
Lynch School of Education, Boston College, Chestnut Hill, MA, USA
e-mail: curt.dudley-marling@bc.edu



D. M. Allen, J. W. Howell (eds.), *Groupthink in Science*,
https://doi.org/10.1007/978-3-030-36822-7_17

based on group membership ignores the reality that the profiles of individuals frequently do not conform to group norms as I show below. The common practice of using data derived from norm-based research to make claims about individual students is an instance of the ecological fallacy that "admonishes us against making inferences about specific individuals based on aggregate data collected from the group to which those individuals belong" (Hlebowitsh, 2012, p. 2).

The normal curve—as applied to the behavior, traits, and abilities of humans—is a myth (Dudley-Marling & Gurn, 2010), an example of scientific groupthink that distorts the meaning of educational research, leading to practices that fail to meet the needs of individuals or subgroups of students whose profiles depart from group norms (e.g., kindergarteners, three-year olds, etc.). In this chapter, I critique the use of the normal curve as a foundation for educational research. I begin by briefly reviewing the research evidence showing that human behavior does not, in fact, distribute normally. This is followed by a discussion of how the ideology of the normal curve distorts educational research and practice. For instance, the use of "the norm" as a reference point for the behavior of individuals creates a vehicle of exclusion for students situated outside the boundaries of "normal" by conflating human differences with deviance. Moreover, the use of the norm as a proxy for group behavior effaces individual differences, obscuring a fundamental insight of the disability studies movement: it is normal to be different. Overall, modern schooling is saturated with the ideology of the normal curve which, by serving the mythical *normal* or *average* child, often meets the needs of no one in particular. Finally, this chapter considers the possibilities of an alternate lens for viewing human behavior that acknowledges the natural variability within groups of people (i.e., "it's normal to be different") as a foundation for organizing schools and conducting educational research.

## Humans Are Not *Normal*

Herrnstein and Murray (1994), in their controversial text, *The Bell Curve*, described the normal curve as "one of nature's more remarkable uniformities" (p. 557). This perspective is widely shared by social scientists, educators, and the general public. As it turns out, however, a substantial body of evidence indicates that the normal curve is a poor representation of social reality that has led to "misguided educational theories, inferences, policies, and practices" (Walberg, Strykowski, Rovai, & Hung, 1984, p. 88).

Sir Francis Galton, one of the first people to advocate the use of the normal curve as a model of human diversity, also provided one of the earliest challenges to the universality of the normal curve. When Galton set out to gather a variety of empirical data to demonstrate the utility of the normal curve he found that, contrary to his expectations, the data for human traits like height, weight, strength, and eyesight failed to produce perfect normal distributions (Micceri, 1989). Similarly, Karl Pearson, a pioneer of modern statistical methods, concluded that, based on his own

observations, a wide-range of phenomena—many cited as textbook examples of normality—did not produce normal distributions (Micceri, 1989). David Wechsler (1935) and Lee Cronbach (1970), major figures in the history of psychological assessment, also cautioned that psychological phenomena do not inherently distribute normally (Fashing & Goertzel, 1981). Geary (1947) went even further, recommending that all statistics textbooks begin with the statement, "Normality is a myth; there never was, and never will be, a normal distribution" (p.241). Although Geary (1947) conceded this statement was a bit of hyperbole, he argued that researchers should never take normality for granted. Indeed, over time, researchers have identified numerous examples of what Bradley (1968) called "bizarre distributions" of human behavior that depart substantially from a normal, bell-shaped distribution.

Despite these challenges to the normal curve as a representation of human behavior, the normal curve continues to exert a powerful influence on educational researchers and practitioners and social scientists more generally (Micceri, 1989). It may be that these individuals have been unduly influenced by the assumption that objective, well-designed achievement and ability tests *necessarily* produce normal distributions that are presumed to be representative of human behavior. Educators may assume, for example, that learning outcomes are normally distributed because achievement scores are presumed to distribute normally. However, achievement tests are "by tradition, custom, or conscious purpose . . . designed to produce such manifest distributions and are not necessarily indicative of the underlying latent [normal] distributions" (Walberg et al., 1984. p. 88). Moreover, the tendency of achievement and ability test data to distribute normally is, to some degree, "simply a mathematical and statistical effect" (Sartori, 2006, p. 415). Standardized educational tests, for example, rely on summated scaling techniques by which persons taking tests attempt to answer a large number of items and receive total scores corresponding to the number of items they answer correctly. This type of measurement has an inherent bias towards a normal distribution in that it is essentially an averaging process, and the central limit theorem indicates that distributions of means tend to be normally distributed (Fashing & Goertzel, 1981; Sartori, 2006). In other words, the average of averages tends to produce normal distributions even if the variables being measured do not distribute normally.

Even given the theoretical bias of objective tests toward normal distributions there is empirical evidence indicating that actual test scores "are seldom normally distributed" (Nunnally, 1978, p. 160). Micceri (1989), for example, examined the distributional characteristics of 440 large-sample achievement and psychometric measures obtained from journal articles, research studies, and national, state, and district tests. Major sources of test data included the California Achievement Test, the Comprehensive Test of Basic Skills, Stanford Reading Tests, Scholastic Aptitude Test (SAT), and the Graduate Record Exam (GRE). In all, Micceri's sample included 46 different test sources and 89 different populations. His analysis indicated that all 440 distributions he examined were "significantly non-normal" (p. 156). It seems that even educational tests designed to produce normal distributions do not necessarily produce such distributions in practice.

The evidence strongly indicates that only truly random events, like the throw of the dice or the flip of a coin, produce normal distributions. Human behaviors are always socially and culturally mediated and, therefore, never occur randomly, a conclusion supported by an overwhelming body of theory and research. Yet the myth of the normal curve as a model of human behavior continues to exert a powerful influence on theory and practice in education and the social sciences, an instance of scientific groupthink that misrepresents the human experience.

In the following sections I consider how the expectation that human behaviors distribute along the lines of a normal curve misleads educational researchers and practitioners.

## How the Ideology of the Normal Curve Distorts Educational Research, Theory, and Practice

The expectation that human behaviors tend to distribute along the lines of bell-shaped, normal curve corrupts how educational researchers interpret their data and how policy makers and practitioners make use of these data. In the sections below, I consider how the myth of the normal curve subverts educators' understanding and use of data from research based on both descriptive and inferential statistics.

### *The Meaning of "Average"*

Educational researchers and policy makers—and even the general public—find means (or averages) useful for describing student characteristics, including the academic performance of various groups and subgroups as well as trends in student achievement over time. For example, student achievement test data by school, school district—or even state or country—are routinely offered up as rough estimates of how well students are achieving in various jurisdictions. Further, disaggregating achievement test data by race or SES over a span of years is often used as a measure of how well schools are addressing historic inequities that have plagued American education and society more generally.

The utility of statistical averages as general indicators of student performance within and across various jurisdictions or within particular groups and subgroups is, however, dependent on the degree to which the mean is a reasonable proxy for the performance of particular groups, that is, a significant proportion of the given population clusters about the mean (the distribution is *normal*). However, the actual distribution of target populations is rarely known by practitioners or policy makers who use these data and, in any case, as the discussion above indicates, human behaviors cannot reasonably be expected to distribute along the lines of a normal curve. Student achievement, for example, is mediated by a host of factors including

the curriculum, class size, teacher experience and expectations, and socioeconomic conditions, none of which is random, a prerequisite for producing a normal, bell-shaped distribution.

Even if achievement test data for groups or jurisdictions did, in fact, distribute along the lines of a bell curve—and, again, this is highly unlikely given the non-randomness of human behavior—the use of the statistical mean to describe the performance of groups of students would still obscure the variation that is always present within any human population. Critics of American education, for example, frequently cite international comparisons to support their claim that U.S. schools are failing to meet the nation's needs. *The Global Report Card*, a website created by the George W. Bush Institute, for instance, states that "the majority of American students are falling behind their international counterparts" and "the consequences to our country could be dramatic" ("The Global Report Card," 2014).The widely cited *Programme for International Student Assessment* (PISA) seems to support this claim. The latest PISA report indicates that U.S. schools rank 25th among OECD countries on various measures of academic achievement (PISA, 2015). While the PISA data certainly invite further scrutiny by policy makers, the relatively poor ranking of U.S. schools, based on statistical averages, masks the considerable variation within and across U.S. schools. For instance, data from the National Assessment of Educational Progress (NAEP), often referred to as the "nation's report card" on the health of American schools, show considerable variability within and across states. For instance, NAEP data indicate that, on average, Massachusetts schools significantly outperform schools in Louisiana in both reading and mathematics achievement (NAEP, 2013). Yet, there are many low-performing schools in Massachusetts and high-performing schools in Louisiana, facts obscured by state averages. Moreover, it is certain that the highest achieving schools in Louisiana outperform the lowest achieving schools in Massachusetts. It may even be the most successful schools in Lousiana outperform the most successful schools in Massachusetts. And, of course, the average performance of particular schools reveals little about the achievement of individual students.

The focus on the average performance of students across nations, states, school districts, and individual schools also masks how factors like poverty affect student achievement. Berliner's (2013) analysis of data from international comparisons, for instance, shows that U.S. students attending schools with relatively low poverty rates do very well compared to their counterparts in other countries. He concludes that, "it is quite clear that America's public school students achieve at high levels when they attend schools that are middle- or upper-middle-class in composition" (p. 7). On the other hand, children and youth attending schools where more than 50% of the children live in poverty do not do nearly as well and students attending schools where at least 75% of the student body is eligible for free and reduced price lunch do even worse. In these schools "academic performance is not merely low: it is embarrassing" (Berliner, 2013, p. 7). Nearly 20% of American children attend these high-poverty schools. But even Berliner's analyses of PISA data can be misleading since high achievers will likely be found in the lowest functioning schools—

and not all children are well served in even the most affluent, highest performing schools.

Some of these problems can be avoided by disaggregating data by groups (SES, for example) or using descriptive statistics that are more sensitive to the variability in any data set (e.g., quartile ranges). But, in the end, the fidelity of descriptive statistics is a function of the underlying distribution and, even then, group averages offer little insight into the behavior, characteristics or abilities of individual students. Writing over 80 years ago in the *Journal of Comparative Psychology*, Knight Dunlap (1935) warned of reporting data on the basis of what he referred to as the "average animal . . . an animal which is entirely mythical" (p. 1). Dunlap observed that in his "list of Great Experiments in Bad Psychology there is one research study in which the average value presented as significant is a value which every person in the experiment conspicuously avoided" (p. 2). Put differently, the statistical average for any particular group of people may apply to no one person in the group. In the context of educational research, the reliance on means to represent groups always risks mischaracterizing individual students, confounding curricular and policy decisions made on the basis of these data.

## *The Meaning of Mean Differences*

Descriptive statistics like averages can be useful for highlighting trends in education or drawing attention to particular issues even if such measures tend to efface individual differences. However, absolute differences between and within groups and subgroups do not necessarily signify meaningful (i.e., non-random) differences. On the NAEP fourth-grade reading rankings for states, for instance, Massachusetts ranks first, Connecticut fourth, and the state of Washington ranks tenth, but it is quite possible that these differences are due to random factors and are, therefore, not meaningful (that is, not *statistically significant*). Nor do absolute differences in mean performance over time permit educational researchers to make strong claims about the efficacy of particular curricular innovations. In order to determine whether mean differences in academic performance between states are "significant" or if targeted instructional interventions are efficacious, educational researchers typically make use inferential statistics.

Consider the example of "best practices," a primary focus of much educational research aimed at identifying effective, evidence-based instructional practices for use in the classroom. The U.S. Department of Education's What Works Clearinghouse, for example, "reviews the existing research on different *programs*, *products*, *practices*, and *policies* in education . . . to provide educators with the information they need to make evidence-based decisions" (i.e., what *works*) (What Works Clearinghouse, 2017). Specifically, the What Works Clearinghouse focuses on "high-quality research," including the use of appropriate statistical analyses that, presumably, permits strong causal claims about the efficacy of particular instructional methods. Typically, this involves some sort of statistical test of mean

differences such as analysis of variance, t-tests, and so on that enable researchers to make assertions about the relative effectiveness of specific interventions assuming researchers have designed studies that eliminate alternative explanations (rival hypotheses) for their results through the use of control groups, group matching, random assignment to groups, and so on.

However, the strongest claim that can be made for even the most carefully designed intervention studies is that particular interventions worked *on average*. In the typical case where specific educational programs or strategies have been found to be effective compared to one or more alternative interventions there is always variability in the data; specifically, no educational intervention has been found to be effective for all students and, indeed, there are always students in comparison groups whose achievement exceeds the mean performance of the experimental groups.

Effect size, a measure educational and other social science researchers routinely compute to determine the meaningfulness of statistically significant differences, is illustrative. Effect size is a useful metric since trivial differences between and within groups can sometimes achieve statistical significance especially with large sample sizes. For instance, an intervention that produced a trivial *improvement* in IQ of just one point could prove to be statistically significant given a sufficiently large sample size. Effect sizes provide a way to gauge the meaningfulness of statistically significant differences and, in the case of an IQ difference of a single point, the effect size would be quite small and, therefore, not meaningful.

Ultimately, effect size, given in standard deviation units, is a measure of variability although it is rarely interpreted that way. An effect size of 0.8, for example, which is considered "large" in social science research (Cohen, 1969), means that, in the theoretical case of a normal distribution, scores for 79% of the control group fall below the mean for the experimental or treatment group. A "large" effect size of 0.8 also means, however, that, again theoretically, 21% of the control group scored higher than the mean for the treatment group. A well-designed study of a reading intervention with a sufficiently large sample size that produced an effect size as large as 0.8 (standard deviation units) would almost certainly qualify as a *best practice*, for example, even though, in this hypothetical case, over 20% of the students in the control condition outperformed the average for the experimental group. Again, this is in the theoretical case where experimental and comparison groups produce normal distributions. In reality, where we can expect non-normal distributions for almost any group of students, the proportion of students for whom the intervention "worked" is, at best, uncertain. What is certain, however, is that even the strongest claims that can be made in support of the most effective educational practices must be qualified with reference to the variability that is always present in any student population, that is, no intervention will work for all of the children all of the time and even the most effective practice may not *work* for a significant proportion of students.

Making assumptions about the potential effectiveness of any practice for individual students based on group means is an instance of an ecological fallacy, an error in reasoning common in how researchers, practitioners, and policy makers

interpret data from educational research (Hlebowitsh, 2012). For example, based on the assumption that best practices work for all or most students, teachers are being directed to teach curricula based on evidence-based practices (Every Student Succeeds Act [ESSA], 2015) with little consideration of students for whom *best practices* are not effective. If students fail to achieve in the presence of best practices the common assumption is that the problem lies in the student who lacks the ability or effort to succeed *normally* (see Dudley-Marling, 2004 for discussion of the social construction of learning failure). Additionally, in the all-too-frequent case where best, evidence-based practices are implemented prescriptively (e.g., Finn, 2009) teachers' professional discretion is circumscribed, making them less effective with students who do not conform to the norm (see Allington, Johnston, & Day, 2002).

Like the descriptive measure of average, tests of mean differences are based on the faulty assumption that human traits and behaviors distribute along the lines of a normal, bell-shaped distribution with most people clustering about the mean. The fetishization of the mean has the effect of masking the range of human differences that are always present in any population of students, perverting educational decision-making in the process.

## Conclusion: It Is Normal to be Different

Recalling the quote at the beginning of this chapter, the idealization of the mean (or average), by obscuring the variability that is always present in any population of students, privileges the "good of the many," students presumed to be more or less average, over "the good of the one," students for whom the mean is a poor representation of their ability or performance. Normative data from even the most comprehensive and well-designed studies routinely mislead educators regarding the needs of individual students who tend not to conform to normative descriptions. This is a case of not being able to see the individual trees for the forest.

The antidote to the "tyranny of the normal curve" is for educators to shift their gaze from measures of normative tendencies to measures of variance. Difference is the norm when it comes to human affairs and this insight ought to change how we conduct and interpret educational research and how we assess and teach students. What about the students for whom "best practices" are not effective, for example? And, more to the point, what about the individual students sitting at desks and tables in elementary and high school classrooms across the country? What do they look like and what sort of instruction do they respond to? Toward this end schools need to create affordances for teachers to provide individual support and direction for students including the assessment of individual student needs and progress monitoring. Recognizing the variability that exists in any group of students also highlights the importance of encouraging teachers to draw on their professional knowledge and experience in support of student learning. It is worth noting that the conceit that there are *best*, research-based practices that should dictate praxis is not limited to

education. The implementation of the best practice service model in medicine and counseling, for example, is widespread with the effect that the professional judgment of physicians and counselors is increasingly devalued.

"Best practices" and other data derived from norm-based research (and assessment) should, at best, be suggestive. Mandating "best practices" because they are research based ignores both the reality of individual needs and the critical importance of teachers' professional judgment. It also effaces the serious limitations of norm-based research practices. In reality the *best* practice is to reject the normal curve as a representation of human behavior. When it comes to the human experience, there is no such thing as a normal curve. It is difference that is the norm.

## References


Allington, R. L., Johnston, P. H., & Day, J. P. (2002). Exemplary fourth-grade teachers. *Language Arts, 79*, 462–466.

Berliner, D. C. (2013). Effects of inequality and poverty vs. teachers and schooling on America's youth. *Teachers College Record, 115*(12). Date accessed: 2/17/2015 http://www.tcrecord.org/library/abstract.asp?contentid=16889

Bradley, J. V. (1968). *Distribution-free statistical tests*. Englewood Cliffs, NJ: Prentice Hall.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York, NY: Academic Press.

Cronbach, J. L. (1970). *The essentials of psychological testing*. New York, NY: Harper & Row.

Dudley-Marling, C. (2004). The social construction of learning disabilities. *Journal of Learning Disabilities, 37*(6), 482–489.

Dudley-Marling, C., & Gurn, A. (2010). Troubling the foundations of special education: Examining the myth of the normal curve. In C. Dudley-Marling & A. Gurn (Eds.), *The myth of the normal curve* (pp. 9–23). New York, NY: Peter Lang.

Dunlap, K. (1935). The average animal. *Journal of Comparative Psychology, 19*, 1–3.

Every Student Succeeds Act (ESSA) of 2015, Pub.L. No. 114-95. (2015). Retrieved from: https://www.gpo.gov/fdsys/pkg/BILLS-114s1177enr/pdf/BILLS-114s1177enr.pdf on January 13, 2016.

Fashing, J., & Goertzel, T. (1981). The myth of the normal curve: A theoretical critique and examination of its role in teaching and research. *Humanity and Society, 5*(1), 14–31.

Finn, P. J. (2009). *Literacy with an attitude: Educating working-class children in their own self-interest*. Albany, NY: SUNY.

Geary, R. C. (1947). Testing for normality. *Biometrika, 34*, 209–242.

George W. Bush Institute. (2014). *The global report card*. Available at globalreportcard.org

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York, NY: Free Press.

Hlebowitsh, P. (2012). When best practices aren't: A Schwabian perspective on teaching. *Journal of Curriculum Studies, 44*(1), 1–12.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156–166.

National Assessment of Educational Progress (NAEP). (2013). *The nation's report card*. Washington, D.C.: National Center for Educational Statistics.

Nimoy, L. (Director), & Nimoy, L. (Writer). (n.d.). *Star trek IV – The voyage home* [Video file].

Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

OECD. (2015). *PISA*: *Results in focus*. Available at https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf

Sartori, R. (2006). The bell curve in psychological research and practice: Myth or reality? *Quality and Quantity, 40*(3), 407–418.

Walberg, H. J., Strykowski, B. F., Rovai, E., & Hung, S. S. (1984). Exceptional performance. *Review of Educational Research, 54*(1), 87–112.

Wechsler, D. (1935). *The range of human abilities*. Baltimore, MD: William & Wilkins.

What Works Clearinghouse. (2017). *Find what works based on the evidence*. https://ies.ed.gov/ncee/wwc/. Accessed 15 Aug 2017.

# Chapter 18
# KETEK


**John M. Norwood, Elizabeth Schriner, and Ah Young Wah**

## Background

The following interaction was not atypical in the 1990s and 2000s during the heyday of the pharmaceutical industry: "Hello, Doctor, I am your pharmaceutical representative for an exciting new drug. It has many, many positive aspects and has minimal to no side effects or drug interactions. May I have your commitment that you will prescribe this product?" That interaction could have, in fact, occurred several times daily in any physician's office. Into that milieu appeared Ketek (generic: telithromycin). The first in a new class of antibiotics, it was expected to be a blockbuster drug and a source of significant profit for its manufacturer. In one of the greatest scandals in the history of the U.S. Food and Drug Administration, enthusiastic approval of the medication led shortly to horror and deceit. Ketek no longer remains on the market today, but reverberations from its stormy background continue.

J. M. Norwood (✉)
Medicine-Infectious Disease, University of Tennessee College of Medicine, Memphis, TN, USA
e-mail: jnorwoo2@uthsc.edu

E. Schriner
Division of Quality of Life (QoLA), St. Jude Children's Research Hospital, Memphis, TN, USA

A. Y. Wah
Regional One Health – East Campus, Memphis, TN, USA



D. M. Allen, J. W. Howell (eds.), *Groupthink in Science*,
https://doi.org/10.1007/978-3-030-36822-7_18

## FDA's Accelerated Approval Process

Historically, the FDA has been extraordinarily conservative after a pregnant Australian woman was given thalidomide in the 1950s and the fetus developed severe birth defects (Hamburg, 2012). For a drug to be approved, it must demonstrate that it is safe and effective in laboratory studies, animal models, and three phases of clinical trials in humans. From this data, detailed and complex statistical analyses can predict outcomes of release into general medical practice. These studies are done by pharmaceutical companies under the oversight of the United States FDA. If a product is successful in these trials, the sponsor may submit a new drug application (NDA). Once approved, the drug enters Phase IV clinical trials and is available for the general medical community; ongoing monitoring is required for years. This protracted review process is designed to allow ample time for investigation of the new agent. During the 1980s and 1990s, however, the human immunodeficiency virus (HIV) epidemic in the United States forced the FDA to speed up the process of drug approval given the urgent need to provide treatment for people with acquired immunodeficiency syndrome (AIDS) (Center for Disease Control, 2011). Given the success of HIV treatment regimens, the FDA accelerated approval of other drugs, which led to the removal of multiple medications from the market after significant problems were discovered post-release. For example, Vioxx, a well-known anti-inflammatory medication, was discontinued for public use after allegations of drug-related heart attacks and strokes surfaced. By the time of its removal, Merck had already sold billions of dollars of the medication worldwide (McIntyre & Evans, 2014).

## Ketek (Generic: Telithromycin) Development

Unfortunately, the worsening resistance of bacterial infections in the late twentieth century has created an ongoing crisis in the availability of safe and effective antibiotic therapy. Erythromycin, a standard treatment for community-acquired pneumonia, a common bacterial infection, was introduced in 1957. It was mostly used for cases of pneumonia if the patient was allergic to penicillin or for cases involving organisms that would not be treatable with penicillin. Second generation macrolides, such as azithromycin and clarithromycin, were developed later by Pfizer and Abbott as similar but better tolerated antibiotics—with fewer drug interactions and a slightly broader spectrum of activity. These antibiotics prevent the development of certain key bacterial proteins by binding to bacterial structures called ribosomes (Fernandes, 2016). Telithromycin, brand name "Ketek," was manufactured by Hoechst Marion Roussel pharmaceuticals (later Sanofi-Aventis) as the first agent in a class of macrolide-like medications, which symbolized an exciting advancement in the war on antibiotic resistance. Ketek was made semi-synthetically by chemically adjusting the structure of erythromycin. Its structure allowed

for binding at two points on the bacterial ribosome instead of one, which helped to prevent the development of bacterial resistance and benefited the effectiveness of the medication (Sanofi-Aventis, 2015).

**Ketek Approval Delays in the United States** Sanofi-Aventis submitted its Ketek new drug application (NDA) to the FDA on February 28, 2000, seeking consent for four indications (community-acquired pneumonia, acute bacterial sinusitis, acute bacterial exacerbation of chronic bronchitis, and pharyngitis), including a claim of effectiveness for drug-resistant Streptococcus pneumoniae. Shortly afterwards, Ketek was approved by the European Medicines Evaluation Agency. In April 2001, the FDA conducted its initial review of Ketek and its Anti-Infective Drugs Advisory Committee voted to deny approval for three of the four indications. The committee requested more safety and efficacy data for these claims, as early evidence in animal models revealed possible liver, heart, and visual side effects. Sanofi-Aventis responded in July 2002 with multiple Phase I studies and three Phase III studies, including the "Randomized, Open-Label, Multicenter Trial of the Safety and Effectiveness of Oral Telithromycin (Ketek) and Amoxicillin/Clavulanic Acid (Augmentin) in Outpatients with Respiratory Tract Infections in Usual Care Settings," also known as Study 3014 (Von Eschenbach, 2007). More than 1800 physicians enlisted in Study 3014, many of them new to clinical investigation. For each patient the provider enrolled, he or she earned up to $400. By the end of the recruitment period, more than 24,000 patients had enrolled (McGoey, 2012).

Here the details of Ketek's background grow murky, and many of the resulting lawsuits and congressional hearings focused their investigations on the events surrounding Study 3014. During these investigations, several healthcare providers and clinic personnel received punishments ranging from lost licenses and fines to prison time. There are still concerns about the level of involvement of the "big fish"—Sanofi-Aventis and the FDA. Did the pharmaceutical company submit fraudulent data knowingly? What prevented the FDA from effectively functioning during this process? Many lives were destroyed by Ketek, and so much of what happened next may have been preventable.

**Ketek Enters the Market** David Ross was one of the FDA physicians who reviewed Ketek's NDA in 2000 and denied approval pending further data. He reported receiving Study 3014, amongst other materials, in July 2002, and then attended a second federal advisory committee in January 2003 to discuss its findings. Around this time, a handful of FDA employees became aware of issues regarding Study 3014's data integrity. However, according to the testimony of FDA Commissioner Dr. Andrew von Eschenbach, inspections had only occurred at three of Study 3014's 1800 sites at the time of the second advisory committee's review. Small pockets of poorly-run clinical sites are not unusual. Therefore, Dr. Ross and the other members of the second advisory committee were not notified of Study 3014's integrity issues. In his later statement to the House Committee on Energy and Commerce, Dr. von Eschenbach (2007) defended this action:

> To avoid compromising any ongoing investigation, it is Agency policy not to publicly disclose even the existence of a pending investigation. Therefore, we could not discuss the data integrity issues of Study 3014 at the public Advisory Committee meeting. However, we also believed, based on the best information available to us, that the concerns applied to only one site out of more than 1800. It is not unusual for data from some sites to be eliminated from a study but to accept data from the other sites. At the time, there was less information about the other sites under investigation.

Unaware of Study 3014's faults, the committee voted 11-1 in favor of Ketek approval. Two weeks later, upon conclusion of its audits involving the first three clinical sites it investigated, the FDA issued an "approvable letter" to the drug manufacturer, which noted unresolved data integrity issues associated with Study 3014 and concerns about incomplete foreign safety data (von Eschenbach, 2007). When the advisory committee convened in March 2003 to discuss other matters, the FDA administrators briefly mentioned that an approvable letter had been issued to Sanofi-Aventis requesting "more information about data from Europe and Latin America" and that final approval also depended on open "inspectional issues" from Study 3014 (von Eschenbach, 2007). The manufacturer responded to the approvable letter in October 2003. As Dr. von Eschenbach (2007) described:

> The October 2003 submission addressed issues of Study 3014 and included post-marketing reports for spontaneous adverse events for approximately four million prescriptions for patients in other countries where Ketek had already been approved. Upon completing the review of the sponsor's October submission, including the findings from the additional audits of clinical trial sites summarized in a March 2004 memorandum from the Division of Scientific Investigations, *the Agency decided that it could not rely on Study 3014* to support approval of Ketek because of the systemic failure of the sponsor's monitoring of the clinical trial to detect clearly existing data integrity problems. Accordingly, Study 3014 was dropped for consideration in making the decision whether to approve Ketek. The Agency considered data from other clinical trials and the international post-marketing experience to conclude there was adequate evidence of safety.

Thus, on April 1, 2004, Ketek graduated to Phase IV trials and was released for public use. The drug was given three indications: acute bacterial sinusitis, acute exacerbation of chronic bronchitis, and mild to moderate community acquired pneumonia in adults. Ketek's official launch by Sanofi-Aventis advertised it as one of the most important innovations in antibiotic therapy. By 2005, its sales reached $193 million (Mathews, 2006).

## Concerns Emerge Regarding Ketek-Associated Liver Damage

Seven months after its approval (February 2005), Ketek's success received its first blow when 26-year-old construction worker Ramiro Obrajero Pulquero walked into a North Carolina emergency room vomiting blood. Doctors diagnosed him with acute liver failure, but could not explain from where the young man had contracted it. He died three days later. His wife was shocked by the sudden nature of his death: "He was a healthy man, strong, and then suddenly we were watching him slip away"

(Mathews, 2006). The only abnormal event at the time of his admission appeared to be a recent nasal infection, treated with a cutting-edge antibiotic. Purported as an outlier by the FDA and Sanofi-Aventis, Mr. Obrajero's story did not reach the headlines. However, by January 2006, Dr. Kimberly Clay and colleagues from the same North Carolina hospital identified two other cases like Mr. Obrajero's. They submitted their findings for the March 2006 issue of the *Annals of Internal Medicine* (Clay et al., 2006). In an unheard-of move, Dr. Harold Sox, editor of the *Annals*, released the report online two months early. "I can't think of a specific instance where we have published a case report like this early," Dr. Sox said. Dr. John Hanson, one of the study's co-authors, added in an interview (Smith, 2006):

> We were stunned by the fact that we saw three cases in one medical center in a very short period of time. It was startling.

Though both Dr. Sox and Dr. Hanson recognized the possibility of coincidence, they felt compelled to report this finding to the wider medical community. "The sooner doctors know about this, the sooner they can take it into account in deciding whether to use the drug," Dr. Sox argued (Smith, 2006).

Shortly before the article's release, higher-ups in the FDA learned of Dr. Clay's findings and conducted an emergency meeting regarding Ketek's safety. The agency issued a public announcement on January 20, 2006, the same day Dr. Clay's report was published online. Incredibly, the announcement cited safety statistics from Study 3014 (despite being officially considered "unreliable" per Dr. von Eschenbach's statement) and supported the drug's approval (Ross, 2007).

**FDA Internal Debate over Ketek Safety** Critics of the Ketek scandal highlighted the inconsistencies in the FDA's position, which appeared to be less in the public's best interest and more in the interest of procuring revenue for the pharmaceutical company. At least four FDA officials—Dr. David Graham (who issued the earliest warnings about Vioxx, too), Dr. Charles Cooper, Dr. David Ross, and Dr. Rosemary Johann-Liang—provided emails and other statements expressing concerns over Ketek's safety to the *New York Times* in June 2006. Referring to Ketek's approval, Dr. Graham wrote: "It's as if every principle governing the review and approval of new drugs was abandoned or suspended where [Ketek] is concerned" (Harris, 2006). He continued:

> The FDA views industry as its client, and that's the only explanation here. The agency saw that it needed to align its interests with the company's, and the company's interest was 'get this drug approved.'

Dr. Cooper added concerns over the FDA's gratuitously forgiving relationship with Sanofi-Aventis: "Given [the company's] track record in which they have proven themselves to be nontrustworthy, [...] we have to consider the possibility that [the staff at Sanofi-Aventis] are intentionally doing a poor job of collecting the postmarketing data to protect their drug sales" (Harris, 2006).

Appalled at the "very serious" problems revolving around Ketek, Sen. Charles Grassley (R-IA) (Harris, 2006) noted later:

> It's no surprise to learn that the F.D.A. didn't listen to Dr. Graham [and Dr. Cooper] on the dangers of Ketek. The F.D.A. has made it their business to discredit [those] who aren't willing to cater to the drug companies.

Grassley told NPR interviewers: "There's got to be respect for the scientific process; and dissident scientists, that have a point of view that might not be the party line, have to be respected" (Silberner, 2006).

In response to the furor from its internal debate and Dr. Clay's article, the FDA did send out a "Dear Doctor" letter about Ketek-associated liver toxicity in June 2006. Over the next year, Ketek prescriptions plummeted and the FDA issued increasingly stronger warnings (Edwards, 2011). Details surrounding the FDA's growing repudiation of Ketek differ radically depending on who tells the story. To some, not only are the discredited clinical sites responsible for Ketek's subsequent fatalities, but the FDA and Sanofi-Aventis should have been held accountable, too. In February 2007, congress decided it would help clear up the matter. As hearing after hearing concluded, the seriousness of the situation was obvious. People were dying, and something had to be done.

**David Ross and NEJM April 2007** In the *New England Journal of Medicine'*s April 2007 issue, David Ross, a former medical officer at the FDA and now a director for the Department of Veterans Affairs, published his perspective on the events surrounding Ketek's fall from grace. In it, he not only claimed to have evidence of fraudulent and ineffective clinical trials, he also indicated that the FDA had been aware of it since before Ketek's approval (Ross, 2007). He reported that FDA managers, in cahoots with Sanofi-Aventis personnel, were negligent in presenting Study 3014 to the second federal advisory committee without mentioning that the study's integrity was under criminal investigation. Specifically, Ross provided a timeline of misconduct and evidence of seemingly purposeful obfuscation, including e-mails and other internal pressure tactics in both the FDA and Sanofi-Aventis. He recalled a meeting with Dr. von Eschenbach in which the commissioner compared the FDA to a football team and threatened to "trade" any players that discussed Ketek's issues outside of the agency (Harris, 2007).

**FDA Response April 2007** In the same *New England Journal of Medicine* issue, several key FDA administrators published a response to Ross' accusations. "Although the FDA did not rely on Study 3014 to support approval, we reviewed the study for safety findings that would have counted 'against the drug,' as is consistent with good review practice," noted Dr. Janice Soreth (Soreth, Cox, Kweder, Jenkins, & Galson, 2007). These administrators, like von Eschenbach, defended the public announcement made in January 2006. Dr. John Jenkins, director of the FDA's office of new drugs, told interviewers that the rate of liver-related problems looked "not all that different than we would see for other antibiotics for similar infections" (Mathews, 2006).

The overwhelming public response to Ross' article, however, was one of concern and mistrust. It was eventually revealed that five of the six authors of the FDA

rebuttal received consulting fees from Sanofi-Aventis, according to a Wall Street Journal exposé. The sixth was an Aventis employee at the time of Study 3014″ (Mathews, 2006). It is too easy to imagine that such connections could have created bias; yet, whatever their motives, key FDA personnel continued to support Ketek.

**FDA Investigation** With such disagreement between Dr. Ross and the 2007 FDA rebuttal letter, what is the real story? Investigations by the agency and Sanofi-Aventis uncovered obvious "bad guys"—i.e. clinical sites that fabricated patient data and forged paperwork—who were prosecuted and fined or imprisoned. Yet, how could Sanofi-Aventis be unaware of such incredible breaches of research protocol? How could FDA administrators not realize the magnitude of the brewing crisis? Are the FDA's critics correct and the agency is now more interested in promoting the interests of Big Pharma instead of public health?

To ensure compliance with the Code of Federal Regulations (CFR), FDA employees conduct intermittent inspections of clinical trial locations. At the conclusion of these inspections, the FDA issues Form 483, which itemizes observations or areas that need to be addressed. Once higher FDA officials receive Form 483, they may send the clinical site a Warning Letter if they deem the observations serious violations. If the site supervisor's response to this letter is inadequate, the FDA opens an official investigation and sends a Notice of Initiation of Disqualification Proceedings and Opportunity to Explain (NIDPOE). When discrepancies involving Study 3014 arose, the FDA sent out these letters to several individuals, including Dr. Keith Pierce and Dr. Maria Anne Kirkman-Campbell.

Dr. Pierce's NIDPOE is an excellent example of the violations involving Study 3014. In it, he is charged with "repeatedly and deliberately submitting false information to the sponsor in a required report" (NIDPOE issued to Pierce, 2010). Specifically, his FDA observer noted the following issues involving the radiologist who was supposed to determine patient eligibility for the study:

1. The signatures on the 'radiologist interpretation worksheet' were forged, unbeknownst to the radiologist.
2. Several patients enrolled in the study did not qualify according to a comparison of the radiologist's initial report (which often showed findings like normal sinuses) and the 'radiologist interpretation worksheet' (which, for the same patient, listed mucosal thickening and even 'total sinus opacity' instead).
3. The radiologist reported potential for bias—he was asked to 'reread' some of the x-rays with the clinical investigator standing over his shoulder.

In addition, a chart review for some of the enrolled subjects revealed that their cases deviated from the study protocol. For instance, two patients received Rocephin and ciprofloxacin (both antibiotics), which were specifically prohibited during the trial duration. Another violation included failure to maintain accurate patient histories and medications (NIDPOE issued to Pierce, 2010).

**Cisneros and Kirkman-Campbell** Dr. Kirkman-Campbell's NIDPOE revealed similar, if not more extensive, deceit. In many ways, she became the face of the

Ketek scandal. Ann Marie Cisneros, a compliance officer from the clinical trial company Pharmaceutical Products Development (PPD) who monitored clinical sites for Sanofi-Aventis, testified about Study 3014's flaws and her investigation of Dr. Kirkman-Campbell before a congressional subcommittee on February 13, 2007. There were warning signs of fraud even before Cisneros personally visited Kirkman-Campbell's office, Cisneros admitted. For instance, the practice had enrolled over 400 patients, an enormous number considering it was located in Gadsden, Alabama (with a population just north of 36,000). No participants withdrew from the trial. Kirkman-Campbell's entire staff and most of her family were part of these highly-dedicated participants. Cisneros also unearthed consent forms that appeared to be forgeries (patients were enrolled at times and on days the office was closed) and patients with no history of conditions relevant to the study would suddenly manifest qualifying symptoms. She e-mailed her findings to her superiors at PPD and Sanofi-Aventis in 2002 (McGoey, 2012). Less than two years later, *the same year Ketek was approved for public use*, Kirkman-Campbell was sentenced to 57 months in prison and ordered to pay $925,000 in restitution to Sanofi-Aventis NIDPOE issued to Kirkman-Campbell, 2006). Again, critics question how such a gigantic misstep was possible—how it was that the FDA actively participated in a criminal investigation involving Ketek's safety data and still permitted its release to the general public.

When she testified before congress, Cisneros was stalwart about the deliberate obfuscation of falsified findings: "Mr. Chairman, I knew it. PPD knew it. And Aventis knew it" (House of Representatives, 2008). Douglas Loveland, one of the FDA criminal investigators who was assigned to the Ketek trial, supported Cisneros' statement but explained why Sanofi-Aventis could never be legally responsible for what happened. At the time of Cisneros' investigation, Sanofi-Aventis did respond to Cisneros' concerns, albeit ineffectively. If they had not, the cover-up could have clearly been prosecuted. However, the pharmaceutical company had records of the actions it took after Cisneros' investigation. Incredibly, Loveland explained, math was to blame for the company's failure to identify imminent catastrophe (House of Representatives, 2008):

> When you get into a traffic accident, you call a traffic cop. [Aventis] came in and they said, 'we have indicators of fraud,' and they called a mathematician. A mathematician didn't know what fraud looked like, and he couldn't identify it. He looked at all the data, couldn't figure out a rule to apply to the data set, came back and said, 'I don't see fraud.' They took that to convince themselves that two of the most serious allegations raised by Ms. Cisneros and by other PPD folks weren't indicators of fraud.

The next mistake Sanofi-Aventis made, Loveland testified, was to issue a "blizzard" of memos to the clinical sites involved in fraud. These memos were meant to address the glaringly obvious patterns of falsified information—the convenient diagnoses, inadequate histories, crossed-out or white-out forms, etc. After signing these memos, the clinical sites were considered "rehabilitated" and their cooperation was forwarded to the relevant oversight agencies. When questioned about their fraud-detecting processes, Sanofi-Aventis agreed its mechanisms were imperfect

(House of Representatives, 2008). Sloppiness, Loveland contended, led to the same flawed decision-making process as intentional fraud: a pharmaceutical company rushed their data (and any investigations of the merits of this information) and patients died. Fortunately for Sanofi-Aventis (and unfortunately for Loveland and his colleagues), the legal system does not equate the two missteps. The pharmaceutical company was acquitted of wrongdoing.

**FDA Involvement and Response** In 2011, a judge upheld Sanofi-Aventis' legal immunity (Edwards, 2011). The FDA, on the other hand, descended into a bitter civil war. After Dr. von Eschenbach's "teamwork" analogy, Dr. Ross and several of his colleagues left the agency. "Without significant changes to our drug safety system and FDA, we are certain to see more Keteks," Ross argued at his congressional hearing (House of Representatives, 2007). David Graham made a similar statement when his testimony helped force Merck to withdraw Vioxx. Graham supported Ross' brutal evaluation of the FDA's problems, as did Rep. Bart Stupak (D-MI): "One must ask, if the FDA is not protecting its client, the American people, whose interest is being protected?" (Richwine, 2007).

In his testimony, Dr. Ross alleged that Dr. von Eschenbach made at least 11 false statements to the House Oversight and Investigation subcommittee (House of Representatives, 2007). Though the FDA refuted these allegations in a letter to congress, Dr. von Eschenbach resigned in 2009 (Mundy, 2008). Despite the motivation for improvement one assumes these events would inspire, an even larger scandal involving the FDA surfaced a year later. A former employee for Cetero Research, a firm that conducted pharmaceutical trials internationally, reported record tampering and falsification of test data. From April 2005 to August 2009, Cetero participated in 1400 drug trials, all of which were suspect. At least 100 drugs had been approved based on these studies. Even today, the FDA refuses to release the names of these medications and many of them are still consumed regularly by patients worldwide (ProPublica, 2013).

## Conclusion

So there it is—a story of murder and deceit. From the time of its approval in 2004 and David Ross' article in 2007, Ketek was prescribed over five million times. Over the past decade, numerous lawsuits have arisen, alleging negligent misrepresentation, defectively designing a medication, failure to warn consumers, deceptive advertising, and more. Questions remain—why was Ketek allowed to remain on the market, what role did Sanofi-Aventis and the FDA play in obfuscating the study flaws and safety data, and will the improvements made after the scandal be effective in preventing similar catastrophes?

One thing is clear: bad science has far-reaching consequences for patients, physicians, pharmaceutical companies, and researchers. Physicians rely on the FDA and pharmaceutical companies to provide evidence of efficacy and safety. Prescribing

habits depend heavily on this data. As mentioned earlier, the Ketek scandal destroyed many lives. Healthcare providers unwittingly gave patients medication that killed them. Consumers trusted the FDA to protect them from dangerous products, not just ineffective ones. Everyone involved in Ketek's development, from the FDA to Sanofi-Aventis, underwent intense scrutiny. Several lost jobs or confidence in their employers. Some were even imprisoned.

However, experts cannot agree whether more stringent review processes and clinical trial guidelines will help or hurt the situation. The Ketek case represents a constant source of tension in medical research between new therapies and safety issues. In his book, *Antibiotics: The Perfect Storm*, Dr. David Shlaes (2010) explains:

> The science of discovering new antibiotics is exceedingly challenging and the economics of antibiotics are becoming less and less favorable. The regulatory agencies like the FDA are contributing to the problem with a constant barrage of clinical trial requirements that make it harder, slower and more costly to develop antibiotics. The pharmaceutical industry, under extraordinary financial pressures, is consolidating at historic rates leaving fewer and fewer large companies standing. The antibiotic market is not as promising as markets for treatment of chronic diseases like high cholesterol or chronic depression or high blood pressure. For those diseases which we cannot cure, the drugs must be taken for long periods of time, frequently for a lifetime. Antibiotics, which actually cure disease, are only taken for days or weeks.

Pharmaceutical companies must recuperate their costs for developing their products, and the development process is extremely expensive. Dr. Shlaes argues (and many healthcare providers and researchers agree) that making the clinical trial process even more complex could dry up the pharmaceutical pipeline for antibiotics. Ultimately, patients and providers may have to decide whether the benefits of these medications outweigh the risks, including the sometimes fatal outcomes surrounding innovation. To Mrs. Obrajero, David Ross, Ann Marie Cisneros, and the numerous other individuals whose lives were forever altered by Ketek, this choice may not seem like much of a choice at all.

The timeline of the rise and fall of Ketek:

- *February 2000*: Aventis submits NDA for Ketek, the first ketolide antibiotic.
- *June 2001*: The FDA declines to approve Ketek for certain indications and requests more safety and efficacy data.
- *October 2001*: Sanofi-Aventis begins enrolling patients in Study 3014. By January 2002, Dr. Marie Anne Kirkman Campbell has already recruited 287 patients.
- *February 2002*: Sanofi-Aventis manager Nadine Grethe gets an email from Pharmaceutical Products Development, which coordinated the clinical trial. The e-mail warns of potential fraudulent activity at Dr. Campbell's location.
- *July 2002*: Sanofi-Aventis submits the completed results of Study 3014 to the FDA, including 407 patients from Dr. Campbell's location.
- *October 2002*: An FDA inspector visits Dr. Campbell's site and notes several protocol violations. Shortly afterwards, inspectors visit Dr. Carl Lange in Illinois and Dr. Egisto Salerno in San Diego. These three providers enrolled the greatest number of patients and all had major safety issues in their data.

- *January 2003*: Dr. David Ross and the second advisory committee meet to consider the Study 3014 data, unaware of the ongoing fraud investigation. The panel votes 11–1 to approve Ketek. The FDA issues an "approvable letter."
- *April 2003*: Dr. Campbell is indicted for fraud and sentenced to 57 months in prison.
- *April 2004*: The FDA approves Ketek, officially not relying on Study 3014 for safety data.
- *February 2005*: Ramiro Obrajero Pulquero dies from Ketek-associated liver failure.
- *January 2006*: Dr. Kimberly Clay of the Carolinas Medical Center publishes about Mr. Obrajero and other possible liver complications in the *Annals of Internal Medicine*. The same day, the FDA issues a public safety announcement citing safety data from Study 3014.
- *April 2006*: The FDA has received 110 reports of adverse events associated with Ketek, including 23 cases of acute liver injury, 12 cases of liver failure, and four deaths, as well as blurred vision and other problems.
- *June 2006*: Four FDA safety investigators express their concerns over Ketek in the *New York Times*. Eventually, the FDA agrees to send out a "Dear Doctor" letter to alert providers about possible liver injury in cases involving Ketek.
- *February 2007*: One day before a congressional hearing on its handling of Ketek, the FDA finally issues a black box warning, the strongest type of safety guidance, for the antibiotic.
- *April 2007*: David Ross publishes his perspective on the FDA's involvement in the Ketek scandal in the *New England Journal of Medicine.* The FDA publishes a response letter in the same issue (Fig. 18.1).

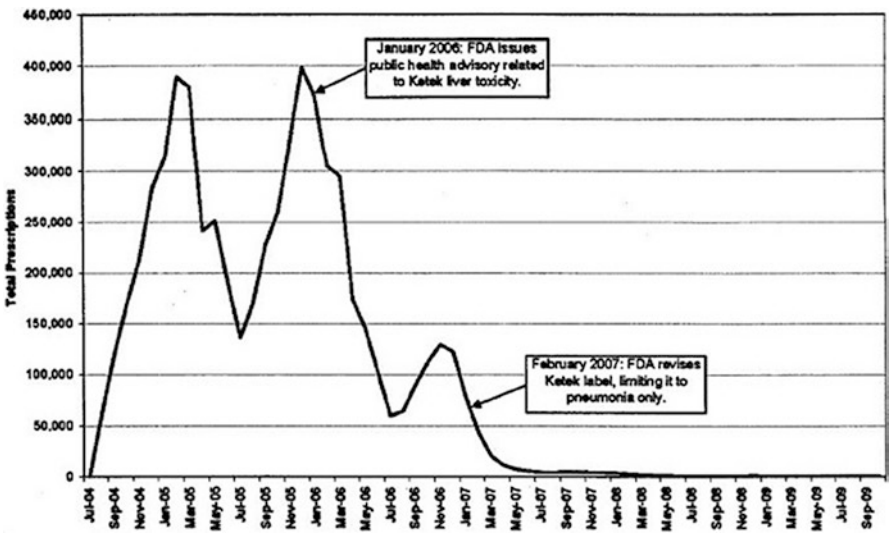


**Fig. 18.1** Number of Ketek prescriptions over time. Edwards (2011)

## References


Centers for Disease Control and Prevention. (2011, January 10). *HIV and AIDS timeline*. Retrieved from https://npin.cdc.gov/pages/hiv-and-aids-timeline

Clay, K. D., Hanson, J. S., Pope, S. D., Rissmiller, R. W., Purdum, P. P., & Banks, P. M. (2006, March 21). Brief communication: Severe hepatotoxicity of Telithromycin: Three case reports and literature review. *Annals of Internal Medicine, 144*(6), 415–420. https://doi.org/10.7326/0003-4819-144-6-200503210-00121

Edwards, J. (2011, February 18). FDA Approved an Antibiotic Based on Bogus Data – But the Courts Don't Care. *CBS News*. Retrieved from https://www.cbsnews.com/news/fda-approved-an-antibiotic-based-on-bogus-data-but-the-courts-dont-care/

Fernandes, P. (2016, August 22). The solithromycin journey – It is all in the chemistry. *Bioorganic & Medicinal Chemistry, 24*(24), 6420–6428. https://doi.org/10.1016/j.bmc.2016.08.035

Hamburg, M. (2012, February 7). *50 years after thalidomide – Why regulation matters*. FDA. Retrieved from https://blogs.fda.gov/fdavoice/index.php/2012/02/50-years-after-thalidomide-why-regulation-matters/

Harris, G. (2006, July 19). Approval of antibiotic worried safety officials. *New York Times*. Retrieved from https://www.nytimes.com/2006/07/19/health/19fda.html

Harris, G. (2007, June 11). Potentially incompatible goals at FDA. *New York Times*. Retrieved from https://www.nytimes.com/2007/06/11/washington/11fda.html

House of Representatives, 110th Cong. (2007). *The adequacy of FDA to assure the safety of the nation's drug supply. Hearings before the subcommittee on oversight and investigations of the committee on energy and commerce*.

House of Representatives, 110th Cong. (2008). *Ketek study fraud: What did Aventis know. Hearings before the subcommittee on oversight and investigations of the committee on energy and commerce*.

Mathews, A. W. (2006, May 1). Fraud, errors taint key study of widely used Sanofi drug. *Wall Street Journal*. Retrieved from https://www.wsj.com/articles/SB114644463095840108

McGoey, L. (2012). The logic of strategic ignorance. *The British Journal of Sociology, 63*(3), 553–576. https://doi.org/10.1111/j.1468-4446.2012.01424.x.

McIntyre, W. F., & Evans, G. (2014). The Vioxx® legacy: Enduring lessons from the not so distant past. *Cardiology Journal, 21*(2), 203–205. https://doi.org/10.5603/CJ.2014.0029

Mundy, A. (2008, Dec 15). FDA Commissioner Eschebach to Resign. *Wall Street Journal*. Retrieved from https://www.wsj.com/articles/SB122938610079708825

NIDPOE issued to Kirkman Campbell, Anne. (2006, May 18). FDA. Retrieved from http://www.circare.org/fdawls3/kirkman-campbell_20060518.pdf

NIDPOE issued to Pierce, Keith. (2010, March 17). FDA. Retrieved from https://www.fda.gov/RegulatoryInformation/FOI/ElectronicReadingRoom/ucm206820.htm

ProPublica. (2013, April 15). *FDA let drugs approved on fraudulent research stay on the market*. Retrieved from https://www.propublica.org/article/fda-let-drugs-approved-on-fraudulent-research-stay-on-the-market

Richwine, L. (2007, Feb 13). *FDA reviews urge action to fix drug safety*. Reuters. Retrieved from https://www.reuters.com/article/us-drugs-fda-s-idUSN1322731520070214

Ross, D. B. (2007, April 9). The FDA and the case of Ketek. *The New England Journal of Medicine, 2007*(356), 1601–1604.

Sanofi-Aventis. (2015). *Ketek: Highlights of prescribing information*. Retrieved from https://www.accessdata.fda.gov/drugsatfda_docs/label/2015/021144s019lbl.pdf

Shlaes, D. M. (2010). *Antibiotics: the perfect storm*. Dordrecht/Heidelberg/New York, NY: Springer. https://doi.org/10.1007/978-90-481-9057-7. Retrieved from https://bfosinfo.files.wordpress.com/2010/09/9048190568antibiotics.pdf

Silberner, J. (2006, December 16). *FDA to leave antibiotic Ketek on the market*. NPR. Retrieved from https://www.npr.org/templates/story/story.php?storyId=6635165

Smith, M. (2006, January 20). Three severe liver damage cases linked to ketolide antibiotic." *Medpage Today*. Retrieved from https://www.medpagetoday.com/gastroenterology/generalhepatology/2525

Soreth, J., Cox, E., Kweder, S., Jenkins, J., & Galson, S. (2007, April 19). Ketek – the FDA perspective. *The New England Journal of Medicine, 356*(16), 1675–1676.

Von Eschenbach, A.C. (2007, March 22). *The future of FDA's drug safety program*. FDA. Retrieved from https://wayback.archive-it.org/7993/20170723044140/https://www.fda.gov/NewsEvents/Testimony/ucm154072.htm

# Chapter 19
# Wildlife Contraception and Political Cuisinarts


**Jay F. Kirkpatrick and John W. Turner Jr.**

## Introduction

New ideas, particularly in the sciences, often elicit strong responses. These run the gamut from rational to dogmatic. This is true even for ideas that have been proven and validated. A variety of organizations, all with sizable constituencies and bearing burdens founded in culture, politics, economics and bureaucracy often feel threatened by new advances because of possible impact on the agendas that serve their own memberships. While their concerns differ, the approach to discredit the new ideas is boringly similar: undermine the idea with purposeful distortions, out of context arguments, irrelevant comparisons and refusal to accept published science. But the results are always the same. At worst the public pushes back from the advances, and at best it becomes ambivalent.

Prior to human intervention, wildlife populations were controlled by the natural processes causing mortality. When animal populations exceeded the carrying capacity of their environment, the environment degraded and resident species died from starvation and disease. Coincidently, the high population density led to a decrease in reproductive success because in animals the age of first breeding was delayed, fewer offspring were produced and juvenile mortality increased.

Historically humans have imposed artificial mortality control upon wild populations through regulated hunting, trapping and poisoning. This was accepted as a normal and essential aspect of human survival. It remains a significant part of human culture and continues to be the primary management tool for some species. However, increasing urbanization, the withdrawal of private lands from the public hunting

J. F. Kirkpatrick (Deceased)
The Sience and Conservation Center, Billings, MT, USA

J. W. Turner Jr. (✉)
Department of Physiology & Pharmacology, University of Toledo College of Medicine, Toledo, OH, USA
e-mail: John.Turner@utoledo.edu



D. M. Allen, J. W. Howell (eds.), *Groupthink in Science*,
https://doi.org/10.1007/978-3-030-36822-7_19

domain, regulatory prohibitions on the use of poisons and trapping, low fur prices and changing public attitudes about lethal wildlife control have reduced the effectiveness of human-induced mortality control as a management tool. We presently face exploding populations of some highly adaptable or protected species but without acceptable management tools for protecting associated environment and animals alike. These events and factors are generally recognized as the impetus behind the emergence of the concept of wildlife fertility control (Asa & Porton, 2005; Rutberg, 2013). The question was how to make it a reality. It seemed that the simplest and least controversial approach to solving the new problems of wildlife management would be the application of existing human hormonal contraceptives to wildlife. Well, nothing is so simple, and the scientists who embarked on this journey to develop new technology and begin applying it might just as well have become involved in the global warming, gay marriage, gun control and universal healthcare issues.

An almost humorous dimension of this is that those same scientists had entered eagerly into this endeavor simply to solve a societal problem. They knew from the start that this endeavor, successful or not, was not a profit-making venture. It was an effort to face the fact that the ever-expanding human population was compressing wildlife space into limited islands of habitat. The options for quelling this rising tide were three: kill, remove (to where?) or slow down reproduction rates. The first was considered unacceptable, the second would run out of space and the third stood alone as logical (Kirkpatrick & Turner, 1985, 1991).

The pursuit of this arcane science, i.e. wildlife contraception, started with hormonal steroids such as used in human birth control, but the key technological breakthrough occurred about 1990 when vaccine-based contraception replaced traditional steroid-based approaches. Very quickly success with a porcine zona pellucida (PZP)–based vaccine was demonstrated in wild horses (Kirkpatrick, Liu, & Turner 1990), white-tailed deer (McShea et al., 1997; Turner, Liu, & Kirkpatrick, 1992), feral burros (Turner, Liu, & Kirkpatrick, 1996), captive exotic species in zoos (Kirkpatrick, Zimmermannn, Kolter, Liu, & Turner, 1995) and a bit later with African elephants (Fayrer-Hosken, Grobler, Van Altena, Kirkpatrick, & Bertschinger, 2000) and bison (Duncan, King, & Kirkpatrick, 2013). To the surprise and chagrin of the researchers, objections were raised. The two species that evoked the loudest cries were wild horses and urban deer.

## Wild Horses

In the case of wild horses, opposition initially arose within wild horse advocacy groups, notably large and politically active organizations. These groups based their objections not on specific scientific arguments, but rather on the wishful thinking that a large, fecund adaptable wild species could be left unmanaged on public lands used for a wide spectrum of purposes, mostly driven by economic interests. A major complicating factor was the passage of the Free-Roaming Wild Horse and Burro Act

of 1971, which imparted almost complete protection to these animals without a hint of effective management.

Predictably, these highly fecund wild horse populations grew from an estimated 17,000 in 1971 to somewhere between 60,000 and 80,000 by the early 1980s. With a myopic view of reality, the interest groups argued for no management under the delusion of self-regulation (which translates into range destruction and starvation), or by predation (despite the absence of effective predation on wild horses in most horse ranges). They also argued that if the economically valuable cattle and sheep were removed from public lands, there would be more habitat for wild horses. However, despite horses representing less than 10% of mouths grazing public land, their lack of economic value made them a ready target.

The opening barrage of opposition to horse population management was not very successful from a legislative standpoint, but litigation and a variety of legal actions became the norm and consequently stalled progress in contraceptive management. The federal government, largely represented by the Bureau of Land Management, installed in 1973 a management system known as Adopt-A-Horse, in which large numbers of horses were captured through helicopter round-up and then removed from the range for adoption of younger ones by the public. Horse injury and mortality were common, and tensions grew between the government and advocacy groups, fueling more litigation. Aside from the questionable humane aspects of this approach, it was neither logistically nor economically successful in keeping up with reproduction on the range (Bartholow, 2008). Lagging adoptions resulted in "surplus" horses that had to be quartered, fed and cared for, and currently there are more than 60,000 wild horses in long-term holding facilities. The annual cost to the taxpayer was more than $75 million in 2013 (De Seve & Boyles-Griffin, 2013) and has continued to rapidly rise to the present.

This management inadequacy combined with encouragement for population control by the National Academy of Sciences provoked some moderate interest by the BLM in fertility control. Between 1977 and the present, the agency has provided varying levels of financial support for the advancement and application of contraceptive technology. However, as time went by, outside demands to expand the application of fertility control became more strident. Intransigence and even opposition to fertility control grew within the agency (National Academy of Science Reports, 1980 and 2013). In fact, despite a clear statement in the 2013 NAS Report that the BLM needed to apply contraception intensively, little change in application rate has occurred to date.

In order to understand the lack of an organized front in moving to a new BLM management paradigm, one has to examine the administrative structure of BLM. Each of the ten western states with wild horses has a state office under the administration of a politically appointed state director. Quite often these directors are appointed for their ability to manage (facilitate) economic uses of public lands (e.g., livestock grazing, mining, energy development and a plethora of recreational uses). Also, each state director is more or less a free agent, and unless their policies are outright illegal, no one in Washington can challenge them. Some state directors

were open-minded and sympathetic to the wild horse plight and some were not, thus there was no coordinated forward movement across the west.

Within each state there are numerous herd management areas (HMAs), each with its own personnel with the responsibility for managing the horses in their HMA. Often consensus on this subject did not exist across HMAs, even within a particular state. Some opposition came from the HMA field managers and was culturally based. A spoken theme delivered to the scientists on numerous occasions by these field managers and crew explained that "we don't do it that way out here; we do it with saddle-horses and ropes." They failed to mention the helicopters that were central to round ups, but their point was clear. This cultural perspective impeded progress.

The Washington and Reno offices of the BLM, which ostensibly oversee all dimensions of the wild horse program, were more or less detached from the realities of the field operations. They too had their conflicts with which to deal. For example, several ranching families throughout the west made millions of dollars annually rounding up wild horses under contract to the BLM. These contracted operations merely mocked the fertility control approach in the early years, but as its application spread, they became vocal opponents. Fertility control was an approach that might cut into the considerable public dollars flowing into their businesses. Some published newsletters complaining about fertility control, and in the Washington (DC) and Reno offices there were those who were sympathetic to keeping the contracted services happy.

As the horse number grew, few state BLM offices paid much attention to the DC/Reno oversight of the program. In 2009 the Washington office sent out a memo to all state offices and HMAs making it clear that when a round-up occurred, any horses to be returned to the range were to be treated with a contraceptive vaccine. A week after the memo went out, an HMA in one western state simply went on the electronic media to declare that this approach did not work, and they would not use it.

Additional cultural and political resistance developed in the central offices. By 2012 one very effective contraceptive vaccine (PZP) was federally approved by EPA and the registration was held by the world's largest animal welfare organization. This particular organization had a history of conflict with the BLM (including instances of litigation), and old wounds were opened. Thus, the DC/Reno offices began to reject fertility control or at least make it difficult to apply, largely because they were disaffected with the organization that held the vaccine registration. One excuse was that BLM did not have the money with which to train BLM personnel to use the vaccine, as required by EPA. Another was that they had problems storing and preparing the vaccine, despite the routine nature of that. A third argument was that too much federal paperwork was required for site-specific permission to use it.

A different version of discontent came from the ranching community. Once again there was little in the way of agreement in this realm. A large segment of the ranching industry, represented primarily by those who used the public lands for grazing, opposed fertility control because they wanted horse removal rather than stewardship. Some in the ranching community were more sympathetic and supported fertility control. A good example of the former was seen with litigation in

2015 by a group of Nevada ranchers who demanded that all horses rounded up be permanently removed, which would indirectly prevent contraceptive use. Various iterations of this approval have been applied on the basis that federal law provides an upper limit for horse numbers on each herd area.

The law requires that a given population exceeding its assigned appropriate management level (AML) be reduced to that level and maintained there or below it. BLM has attempted to do this almost exclusively by removal of horses and has failed overall. The program-wide horse numbers on the range in 2019 are rapidly approaching 90,000 in the face of an agency goal of 35,000. This situation exists despite BLM's own funding of research yielding significant PZP vaccine improvements (e.g., Turner et al., 2008; Turner, Liu, Flanagan, Bynum, & Rutberg, 2002) and the regular pleas of the science community for the past 25 years to incorporate aggressive program-wide vaccine contraception into wild horse management.

A good example of this was published in 2013 in the widely read journal *Science* (Garrott & Oli, 2013). Unfortunately, and despite such appeals, since 2017 the agency has taken a mantra-like stance of "remove to AML, before any contraception." This position creates a quandary, since AML has been unavailable for many HMA's despite the effort. The fact is that a coincident combination approach of some (e.g., 50%) "catch/removal" and some (e.g., 50%) "catch/contraception/release" is the solution supported by data-based modelling. This information in various forms has been provided to the agency since 2012. At this point the cost either way is monumental.

As a retrospective on how damaging culture and politics can be to scientific progress and outcome application, it is noteworthy that the National Park Service began using vaccine contraception as the lone management tool for the wild horse population on Assateague Island National Seashore in 1994 and has successfully continued this form of management to the present (Kirkpatrick & Turner, 2008). While identical vaccine application eventually occurred in select small wild horse populations in the western United States, it required 8 years of regular pestering. It has been highly successful but has required the concerted effort of a few committed BLM field managers and local citizens. In other words, the BLM is not embracing it.

In deference to the agency, many herd areas (HMAs) contain horse numbers in the many hundreds that are not readily accessible by darting. However, the BLM has known since 2004 (because BLM funded the research) that a one-injection vaccine with 1–2 years of effectiveness was available for treating the many gathered mares that were returned to the range, thus preventing thousands of foals. BLM instead treated only as a small percentage, expressing various "reasons" but again not embracing contraception. The long delay in BLM approval and the continued limited acceptance of contraception in DC and Reno again have reflected the power of misinformation and personal, cultural and political bias.

It is worth noting that the relative autonomy at the local and state level of the agency has more enabled status quo rather than progress in horse population management. On the other hand it is not fair to fault the many employees who are doing the best they can in the face of the local realities they deal with, i.e., ranchers, horse advocates, habitat advocates, recreationists, anti-government souls, loonies and the

paperwork and logistics of multi-tasking land stewardship. As with any organization, some folks are dedicated and some are just seeing a job. However, the byzantine nature of the agency does not foster timely progress. As is the case with most giant organizations, bureaucracy is the gun with which the agency shoots itself in the foot.

Going back to the subject of logic regarding wild horse control, it is ironic that with only a few exceptions, the wild horse advocacy groups reversed track by 2010 and began to support fertility control. This reversal was based on their realizing that the only three choices were (1) range destruction/starvation, (2) round up and removal or (3) fertility control. Predation, self-regulation and disappearance of livestock were simply never going to happen. Thus, they embraced fertility control by default. That reversal, while friendly to the fertility control paradigm, only seemed to increase the polarization with the larger BLM (if the advocates like it, we do not!)

The optimistic beginning to a perceived solution for regulating wild horse populations by a small group of scientists trying to find a better and more humane future for innocent animals morphed into a cultural and political nightmare. No one was prepared for the firestorm that came from their efforts, and to date the solution remains within reach but unrealized. In April 2019 a document focused on "A Path Forward for Management of BLM's Wild Horses and Burros" was put forth by a coalition of 12 organizations of varying purpose to provide Congress and BLM with a clear picture of issues, approaches and a long-term view for addressing this crisis. Thus, the effort continues.

## Urban Deer

The controversy surrounding urban deer fertility control is less convoluted than with horses but far more intense. Again, it caught the scientists by surprise. What could possibly be controversial about inhibiting reproduction in urban deer that are eating shrubbery, causing car accidents and damaging the remaining urban woodlands?

The possible application of fertility control for controlling urban deer populations via a contraceptive vaccine was first broached in 1988 at a Princeton conference, and reactions by managing agencies ranged from frowns to amusement. The managing entities consisted of state fish and game agencies, which by law are responsible for wildlife management in their respective states. Soon after the conference, several organizations (including a New Jersey arboretum, a public park in Philadelphia and a group of small communities on Long Island, NY) began lobbying for fertility control. The state agencies sobered a bit and began pushing back.

Initially their arguments against urban deer fertility control centered on a list of hypothetical biological consequences and to a lesser extent on the cultural philosophy that hunting was the only solution (Turner, 1997). Based on these objections, states in which projects were proposed (NJ, PA and NY) simply refused to issue permits to conduct any trials. Subsequently, Turner et al. (Turner et al., 1992) demonstrated that the PZP vaccine (same as used in horses) provided excellent contra-

ceptive efficacy in captive white-tailed deer. Captive studies continued, and researchers requested permission to perform field trials in Metro parks where hunting is prohibited.

By this time, an unspoken undercurrent was developing in state wildlife agencies that contraceptive management of deer living in urban communities and city parks could somehow become a threat to recreational hunting. Driving that concern in part were declining hunting-license sales across the United States and the potential further loss of revenue if deer contraception expanded. While the agency revenue loss would be significant, the potential loss of ancillary revenue related to hunting would be enormous. Hunters buy weapons and gear, stay in motels, put gas into their vehicles, dine in restaurants, purchase ammunition, etc. This concern led the commercial facets of the hunting industry to take a stand against fertility control. The state agencies in turn blurred the lines between urban deer and truly wild deer in the forests. Opposition grew, and urban deer kept eating ornamental shrubbery.

However, states do not have jurisdiction over wildlife on federal lands, so the scientists found several federal field sites for testing the idea of managing urban deer without the need for state approval. The first was a trial at the Smithsonian Institute's Conservation and Research Center in Front Royal, VA. The trial was successful and generated useful data (McShea et al., 1997). The second test occurred in a group of small communities on Fire Island National Seashore (FIIS), in NY. This was a National Park Service (NPS) unit and beyond the legal jurisdiction of the state. As plans progressed, however, the state raised strident objections, all based on "biological" issues. By this time, counterparts in PA and NJ had also refused to allow fertility control to move forward and were beginning to coordinate their objections. It was clear that the issue was a powder keg and that other states were going to join in the effort to prevent urban deer fertility control.

Despite this, the project on FIIS went forward because of the federal classification. The New York Department of Environmental Conservation (DEC), realizing that the project seemed inevitable, threatened the NPS with a lawsuit to stop the project. The NPS, through its regional science office, responded with a terse message that challenged NY to see who really did have authority there. Cooler heads prevailed in Albany and the project went forward. To illustrate the degree of threat seen by state agencies, it is notable that the head of the New Jersey fish and game agency threatened to sue New York for allowing the project to get underway. No action was taken, but it indicated the seriousness with which states viewed fertility control as a threat.

The NPS and even the scientists were also soon informed of a possible lawsuit by a collection of hunting groups on Long Island. While nothing came of that, it signaled the entrance of the larger hunting community into the fray. In the end the project went forward and after 17 years of fertility control the deer population in these communities was reduced by 70% without the removal of a single deer (Naugle, Rutberg, Underwood, Turner, & Liu, 2002; Rutberg, Naugle, Turner, Fraker, & Flanagan, 2013, Rutberg, Naugle, & Verret, 2013; Rutberg & Naugle, 2008). The published data were to become a thorn in the side of all the concerned state agencies. By 1993 the state agencies publicly opposed deer contraception. At

the Third International Conference on Fertility Control in Wildlife (Denver, CO), the agencies showed up in force. They were careful to not emphasize the subject of hunting in city parks, instead focusing on modeling (not data-based) with a bias against contraception.

Despite this opposition, a third major project was born at the National Institute of Standards and Technology (NIST) in Gaithersburg, MD. This one also had a rocky start. NIST is a facility of the U. S. Department of Commerce and once again outside the jurisdiction of the state of Maryland. Maryland Department of Natural Resources (DNC) strongly objected and when NIST refused to allow them to have a hunt on the one-square mile, heavily populated (6000 employees) research facility, the Maryland agency went to the Congressional Sportsman's Caucus. NIST officials and the U. S. Department of Commerce refused to give in and petitioned the U. S. Solicitor General's office for a clarification of the law. The results were predictable; Maryland had no wildlife management authority on the NIST campus. The project went forward and more data were forthcoming and published (Rutberg, Naugle, Thiele, & Liu, 2004).

However, Maryland DNR remained persistent. As the project started, they threatened to ring the facility with agency personnel and shoot any deer leaving the grounds, which are surrounded by heavily traveled highways and residential areas (a suburb of Washington, DC). At that point a local animal welfare organization pointed out that this would make a wonderful media opportunity for the evening news. Maryland backed off temporarily.

Approximately 2 years into the project, Maryland DNR asked for a meeting with the research team and NIST officials and asked if they could conduct a test on the health of the deer. It is worth mentioning that when the project went forward, DNR was invited to participate and take blood samples or make any measurements they deemed valuable. They declined. Many questioned why Maryland DNR waited 2 years to seek permission to kill 50 deer and "assess their health." At that point the research-team veterinarian asked the Director of the Maryland DNR if, when he took his dog to the veterinarian, the dog had to be killed to assess its health. Thereafter no further communications of note occurred between NIST and Maryland DNR.

As might be expected, the researchers developing and testing the PZP vaccine were having their own share of frustration in the face of what seemed illogical resistance to its use. Because of the public notoriety of the deer contraception subject, they experienced many interviews and spoke at numerous public community meetings, stressing their purpose that deer fertility control focus was for parks, preserves and communities where hunting was illegal. A reporter at one of these interviews said he was told state agencies were concerned that fertility control is a threat to hunting. Author Kirkpatrick smiled and after a brief pause said "If those folks think that some guys with dart guns can control state wild deer populations, they must be smoking something really good."

Through the 1990s, the attacks directed at urban deer fertility control by state fish and game agencies were largely based on scientific questions. Chief among these questions and almost identical to ones asked by opponents of PZP for horses were

(1) possible passage of the vaccine through the food chain, (2) possible extension of the breeding season with energetic consequences to the females, and (3) possible genetic effects. In no case were data or evidence of any kind offered to support the concerns. Coincidentally, through the three ongoing projects, an extensive database was generated that answered the questions for deer. The vaccine antigen could not survive the digestive tract. Extension of the breeding season was minimal and did not cause notable energetic loss. In fact, the weight of treated deer improved relative to deer that became pregnant and faced a summer of lactation.

Finally, when compared to the genetic effects of hunting, where the largest and most robust animals were selected against (for their antlers), contraception was a bargain. As numerous biological questions were gradually answered by the ongoing research (Kirkpatrick & Turner, 1995), the demand for the fertility control approach grew in the public sector and public meetings on the subject became more strident and were eventually tinged with hostility as people spoke their views. Unfortunately for all involved, the entrance of animal welfare/protection groups on the side of fertility control led to a deep polarization with the state agencies, which continued to selectively quote and ignore science as it suited their goals (Kirkpatrick & Turner, 1997; Rutberg et al., 1997).

Eventually realizing that the attack on the science was failing, the state agencies turned their attention to regulatory issues. Beginning in 1992 the application of PZP vaccine for deer was regulated by the Center for Veterinary Medicine at the Food and Drug Administration (FDA). Application of PZP was authorized by means of an investigational new animal drug exemption (INAD), the equivalent of INDs issued for the use of unapproved human drugs. Thus, the vaccine had official legal federal authorization. Subsequently a coalition of 16 states that disliked the concept of fertility control lobbied their respective Congressional delegations with the message to get the whole business stopped. However, FDA refused to give in to the political pressure. For states, it was "strike two."

By 2005, the regulation of wildlife contraceptives had been transferred from FDA to the Environmental Protection Agency (EPA). Shortly thereafter a second potential deer contraceptive vaccine was developed by the U. S. Department of Agriculture (USDA). This was a vaccine against gonadotropin-releasing hormone (GnRH) and was named GonaCon®. When USDA applied for registration with EPA, the states descended upon the agency. In the end they could not stop the registration, but they were successful in convincing EPA to place use restrictions on it, e.g., having to capture and tag each deer. The agencies argued that this requirement was to insure that persons harvesting deer would know if it was treated. This was despite the well-established fact that the vaccine was harmless if ingested. However, it did serve its purpose by markedly reducing the practicality of using the vaccine.

In the meantime, two smaller urban deer contraceptive field-research projects were established at the Columbus, OH, Metroparks facility and on Fripp Island, SC. In both cases the respective states approved research permits, which in itself was progress. But after several years of successful application of fertility control and decreases in population growth, both states rescinded their permits on the basis that these sites were actually "managing" deer rather than just doing research. Thus,

the third phase of the states' attack on urban deer fertility control emerged, and it was spectacularly successful. It appeared that the new approach would be to establish state regulations and policies that would prevent fertility control from ever gaining traction. The approach quickly became implemented. State after state established these policies and regulations. Nebraska went so far as to amend its constitution to discourage the use of fertility control.

When GonaCon® was approved by EPA, Pennsylvania was asked to develop a state policy for urban deer fertility control. Carl Roe was then the Executive Director of the Pennsylvania Game Commission and stated publicly: "GonaCon® will never be used by the Game Commission so long as I am director." In 2012 at the Seventh International Conference on Fertility Control in Wildlife, the USDA's John Eiseman provided a list of 17 state policies, of which most were hostile to the concept of fertility control for urban deer management (Eisenman, O'Hare, & Fagerstone, 2013). By 2015 the states had largely won the deer-contraception battle through state regulations/policy.

Despite this success in blocking the application of contraceptives to urban deer management, the states still had to contend with the very impressive successes on the two federal sites, FIIS and NIST. In addition continued research had led to development of a single-injection, multi-year PZP vaccine, which would reduce the access issue and make treatment more practical (Turner et al., 2008). These successes could not be pushed aside or ignored. They kept coming back in the forms of scientific publications, popular media articles and, most importantly, strident public sentiment. Something had to be done to remove this thorn, especially since many communities with deer issues continued to explore fertility control as an option, causing the conflict to fester.

State pressure turned to the National Park Service, the parent agency for FIIS. Strategically this organization had to be reined in because it had two ongoing wild horse fertility control projects and one with wapiti as well as the FIIS project. Even more important was the fact that the NPS was a focal point for many potential fertility control projects. There were additional horses (Mesa Verde NP, Theodore Roosevelt NP), wapiti (Rocky Mountain NP, Point Reyes National Seashore), feral burros (Virgin Islands NP), bison (Yellowstone NP) and mountain goats (Olympic NP) among others, where varying degrees of pressure were being applied for the introduction of fertility control. Even more concerning to the states that were heavily entrenched in opposition to deer contraception were the multitude of potential deer projects (Indiana Dunes, Valley Forge, Gettysburg, Rock Creek Park, in Washington, DC, and several dozen other sites within the NP system).

The irony to this lies in the fact that the NPS has historically been the leading edge for application of fertility control to various wildlife populations. The single largest scientific breakthrough had occurred in an NPS unit (Assateague Island National Seashore) with wild horses, and the application rapidly spread to FIIS for deer, more horses at Cape Lookout National Seashore, wapiti at Point Reyes National Seashore, etc. So, now the states were faced with getting the NPS under control.

The precise strategy, mechanisms and intrigue behind this new effort remain obscure, but the results were soon forthcoming. In 2009, at the urging of several state agencies, the NPS met and established a set of five criteria for deer contraception in NPS units. One criterion was a contraceptive that would have 5 years of efficacy with a single administration. Such a contraceptive did not exist then, and the chances of such a contraceptive being developed are small. Together with the edict that all treated deer have to be ear-tagged (which eliminates remote treatments and increases costs in a significant manner) NPS deer contraception was made virtually impossible. But state pressure on the NPS did not stop there.

The primacy of federal law provides the NPS with all the authority they need to pursue urban/park deer fertility control, yet they deferred to the states. When asked why the NPS did not pursue deer fertility control at Valley Forge where the Pennsylvania Game Commission objected, one regional scientist for the NPS stated "we want to be good neighbors". In 2010, after 17 years of very successful deer control on a five-community block on FIIS, the NPS terminated the project, because "it had to be studied more." Eight years later there is no study and no fertility control on FIIS. And finally, in 2012, at the Seventh International Conference on Fertility Control in Wildlife, the NPS issued a policy statement on wildlife fertility control in NPS units (Wild, Powers, Monello, & Leong, 2013). Two critical items were (1) fertility control methods were considered "more acceptable in non-native species, closed populations, and highly manipulated environments" and (2) "early and active engagement with neighboring state and federal management agencies and public stakeholders is crucial for program success."

Despite this policy, the NPS actively opposed deer fertility control in "highly manipulated environments" such as Rock Creek Park in Washington, DC, and with what they consider to be "non-native species" such as horses in Mesa Verde National Park as well as on other sites. Interestingly the question of whether the horses are a native or non-native species is in itself contentious (Kirkpatrick & Fazio, 2010).

## Conclusion

By 2018, wildlife fertility control was actively being applied to wild horses in more than 35 U.S. sites, including units of the NPS, BLM, U.S. Forest Service, several Indian reservations, a dozen wild horse sanctuaries, the Canadian Province of Alberta and in Hungary and Romania. Nonetheless, only a few of these are actively managing horse numbers in ongoing fashion. The technology has also spread to African elephants, where 20 game parks in South Africa are successfully managing their animals with fertility control and culling is off the Table. A herd of feral sheep in England is also being managed with a contraceptive vaccine. Two different U.S. bison populations have been treated with a contraceptive vaccine, with one realizing zero population growth in a single year. Currently, 4 deer fertility control projects are ongoing, and several have been completed in various communities. More than 200 zoos worldwide are using the same contraceptive technology for the

management of more than 85 species in order to reduce or eliminate "surplus" animals, since disposition is difficult and fraught with controversy.

On the other hand, during this period the NPS sanctioned the culling of deer in Valley Forge, Rock Creek Park and several national historic sites. Yellowstone Park sent between 500 and 900 bison off to slaughter in 2014–15 in order to meet population goals. Mesa Verde National Park refused to even discuss wild horse management by fertility control, while at the same time fencing the horses out from the few available water supplies. Point Reyes National Seashore wrings its hands over a growing and damaging wapiti herd in the face of earlier demonstration that fertility control was a viable option. Hundreds of communities across the U.S. spend hundreds of thousands of dollars to have commercial "sharpshooters" come into their towns and parks and shoot urban deer. In the horse realm the Bureau of Land Management continues to remove and warehouse horses at immense expense to the taxpayer and with associated detriment to quality of life for thousands of horses. Furthermore, the inertia of insignificant use of proven PZP-based contraception continues 2 years after publication of strong evidence that it is an effective long-term contraceptive (5–6 years of infertility across 7 years) and that it really does limit population growth (Rutberg, Grams, Turner, & Hopkin, 2017).

The science of wildlife contraception has been thoroughly vetted within the scientific community through numerous publications in peer-reviewed journals and eight international conferences on the subject. At the field level the actual application of fertility control to free-ranging wildlife is not without its difficulties. The approach is labor intensive compared to other management paradigms, and not all populations will lend themselves to effective treatment and management because of differences in population sizes and habitat. Nevertheless, wildlife fertility control has proven itself a useful management tool. A more detailed discussion of field aspects of wildlife contraception is provided by Turner and Rutberg (2013).

A key consideration for the future of wildlife fertility control is the need for greater crossover of information into the public sector and to Congress about the curative capabilities of wildlife contraception for species in the dilemma of overpopulation. However, even with that accomplishment a crucial obstacle to moving forward is human nature. The desire to defend one view and attack the opposite is hard wired. While some individuals can think their way through to compromise, others cannot or will not. When individuals of the latter case are in positions of control and have decision-making power, ego and defensiveness will rule, and education toward compromise will not readily occur. Therefore, it is important to persist. The long journey continues, sustained in part by concern for pressed species and their environments and in part by those believing that fact and logic will eventually shine through the cloud cover of political agendas, cultural inertia and egocentric bias.

Every spring another cycle of birth plays out for wild horses, deer and many other species. This insurance for species preservation is strong. Across many generations species and habitats will flourish and decline. Human impacts are now figuring heavily into these patterns as part of the cost of human accomplishments.

Perhaps we as a global community can evolve sufficiently to prevent the environmental chaos that can result from continued lack of attention to these patterns. Certainly, focused local attention and action is a realistic goal. However, to accomplish this we must remove the slavery of personal bias and self-serving that derive from cultural and political indoctrinations infused across our own (human) generations. We need to think cleanly and seek the long view. Can we actually accomplish that? Yes, because knowledge and education are great vehicles of science. A positive information/education program focused on compromise and means of resolution can pave the way. Remember, humanity once believed that the world is flat. Some still do.

## References

Asa, C. A., & Porton, I. (Eds.). (2005). *Wildlife contraception: Issues, methods and application*. Baltimore: Johns Hopkins University Press.

Bartholow, J. M. (2008). Economic benefit of fertility control in wild horse populations. *Journal of Wildlife Management, 71*, 2811–2819.

De Seve, C. W., & Boyles-Griffin, S. (2013). An economic model demonstrating the long-term cost benefits of incorporating fertility control into wild horse (*Equus caballus*) management programs on public lands in the United States. *Journal of Zoo and Wildlife Medicine, 44*, 34–37.

Duncan, C. L., King, J. L., & Kirkpatrick, J. F. (2013). Romance without responsibilities: The use of the immunocontraceptive vaccine porcine zona pellucida to manage free-ranging bison on Catalina Island, California, USA. *Journal of Zoo and Wildlife Medicine, 44*, 123–131.

Eisenman, J. D., O'Hare, J. R., & Fagerstone, K. A. (2013). State level approaches to managing the use of contraceptives in wildlife in the United States. *Journal of Zoo and Wildlife Medicine, 44*, 47–51.

Fayrer-Hosken, R. A., Grobler, D., Van Altena, J. J., Kirkpatrick, J. F., & Bertschinger, H. J. (2000). Immunocontraception of free-roaming African elephants. *Nature, 407*(6801), 149.

Garrott, R. A., & Oli, M. K. (2013). A critical crossroad for BLM's wild horse program. *Science, 341*, 847–848.

Kirkpatrick, J. F. (2008). Achieving population goals in long-lived wildlife (Equus caballus) with contraception. *Wildlife Research, 35*, 513–519.

Kirkpatrick, J. F. & Fazio, P. M. (2010). *Wild horses as Native American wildlife*. The Science and Conservation Cntr, Zoo Montana, Billings.

Kirkpatrick, J. F., Liu, I. K. M., & Turner Jr., J. W. (1990). Remotely delivered immunocontraception in feral horses. *Wildlife Society Bulletin, 18*, 326–330.

Kirkpatrick, J. F., & Turner Jr., J. W. (1985). Chemical fertility control and wildlife management. *Bioscience, 35*(8), 485–491.

Kirkpatrick, J. F., & Turner Jr., J. W. (1991). Reversible fertility control in non-domestic animals. *Journal of Zoo and Wildlife Medicine, 22*, 392–408.

Kirkpatrick, J. F., & Turner Jr., J. W. (1995). Urban deer fertility control: Scientific, social and political issues. *Northeast Wildlife, 52*, 103–116.

Kirkpatrick, J. F., & Turner Jr., J. W. (1997). Urban deer contraception: The seven stages of grief. *Wildlife Society Bulletin, 25*, 515–519.

Kirkpatrick, J. F., Zimmermannn, W., Kolter, L., Liu, I. K. M., & Turner Jr., J. W. (1995). Immunocontraception of captive exotic species. I. Przewalskis horse (*Equus przewalskii*) and banteng (*Bos javanicus*). *Zoo Biology, 14*, 403–413.

McShea, W. J., Monfort, S. L., Hakim, S., Kirkpatrick, J. F., Liu, I. K. M., Turner Jr., J. W., et al. (1997). Immunocontraceptive efficacy and the impact of contraception on the reproductive behaviors of white-tailed deer. *Journal of Wildlife Management, 61*, 560–569.

National Academy of Sciences. (1980). *Wild and free-roaming horses and burros: Current knowledge and recommended research*. Washington, D.C.: National Research Council.

Naugle, R.E., Rutberg, A.T., Underwood, H.B., Turner, J.W. Jr., Liu, I.K.M. (2002). Field testing of immunocontraception on white-tailed deer (*Odocoileus virginianus*) on Fire Island National Seashore, New York, USA. *Reproduction*, (Suppl.60):143–153.

Rutberg, A. T. (2013). Managing wildlife with contraception: Why is it taking so long? *Journal of Zoo and Wildlife Medicine, 44*, 538–543.

Rutberg, A. T., Grams, K., Turner Jr., J. W., & Hopkin, H. (2017). Contraceptive efficacy of primary and booster doses of controlled-release PZP in wild horses. *Wildlife Research, 44*, 174–181.

Rutberg, A. T., & Naugle, R. E. (2008). Population-level effects of immunocontraception in white-tailed deer (*Odocoileus virginianus*). *Wildlife Research, 35*, 494–501.

Rutberg, A. T., Naugle, R. E., Thiele, L. A., & Liu, I. K. M. (2004). Effects of immunocontraception on a suburban population of white-tailed deer (*Odocoileus virginianus*). *Biological Conservation, 116*, 243–250.

Rutberg, A. T., Naugle, R. E., Turner Jr., J. W., Fraker, M. A., & Flanagan, D. R. (2013). Field testing of single-administration porcine zona pellucida contraceptive vaccines in white-tailed deer (*Odocoileus virginianus*). *Wildlife Research, 40*, 281–288.

Rutberg, A. T., Naugle, R. E., & Verret, F. (2013). Single treatment porcine zona pellucida immunocontraception associated with reduction of a population of white-tailed deer (*Odocoileus virginianus*). *Journal of Zoo and Wildlife Medicine, 44*, 75–84.

Rutberg, A. T., et al. (1997). Lessons from the urban deer battlefront: Plea for tolerance. *Wildlife Society Bulletin, 25*, 520–523.

Turner Jr., J. W. (1997). Immunocontraception in white-tailed deer. In T. J. Kreeger (Ed.), *Contraception in wildlife management* (pp 147–159). Technical Bulletin 1853. Washington, D.C.: US Dept. Agriculture.

Turner Jr., J. W., Liu, I. K. M., Flanagan, D. R., Bynum, K. S., & Rutberg, A. T. (2002). Porcine zona pellucida (PZP) immunocontraception of wild horses (Equuscaballus) in Nevada: A 10 year study. *Reproduction Supplement, 60*, 177–186.

Turner Jr., J. W., Liu, I. K. M., & Kirkpatrick, J. F. (1992). Remotely delivered immunocontraception in captive white-tailed deer. *Journal of Wildlife Management, 56*(1), 154–157.

Turner Jr., J. W., Liu, I. K. M., & Kirkpatrick, J. F. (1996). Remotely delivered immunocontraception in free-roaming feral burros (*Equus asinus*). *Journal of Reproduction and Fertility, 107*, 31–35.

Turner Jr., J. W., & Rutberg, A. T. (2013). From the pens to the field: Real-world wildlife contraception. *Journal of Zoo and Wildlife Medicine, 44*(4S), S102–S110.

Turner Jr., J. W., Rutberg, A. T., Naugle, R. E., Kaur, M. A., Flanagan, D. R., Bertschinger, H. J., et al. (2008). Controlled-release components of PZP contraceptive vaccine extend duration of infertility. *Wildlife Research, 35*, 555–562.

Wild, M. A., Powers, J. G., Monello, R. J., & Leong, K. (2013). Ungulate fertility control in units of the National Park Service: Future directions. *Journal of Zoo and Wildlife Medicine, 44*, 149 (abstract).

# Chapter 20
# The Influence of Groupthink During the Invention of Stanley Milgram's Obedience Studies


**Nestar Russell**

In 1933, the same year the Nazi regime ascended to power, Stanley Milgram was born into a working class Jewish family in the Bronx in New York City. During his formative years, Milgram was perturbed by the Holocaust. Later he became a social psychologist and obtained a tenure track position at Yale University. During the many Nazi war crime trials, "ordinary" Germans in the docks—like Adolf Eichmann in Israel—typically explained that in participating in the Holocaust they were just following higher orders. This led Milgram to wonder what would happen if he ran a social psychology experiment where ordinary (American) people were ordered to inflict harm on another person. Would they also do as they were told? He designed a basic procedure that tested this question and soon afterwards had his students run the first pilot.

The result from the first trial stunned Milgram—most subjects indeed obeyed orders to inflict what appeared to be intense shocks on an innocent person. Milgram immediately sensed he had captured essential elements of the Holocaust in the laboratory setting. Thereafter he applied for funding to run an official research program so that he could better understand so-called obedience to authority. Milgram's intentions were not entirely honorable—running such an innovative research program could greatly boost his then precarious career prospects and financial security. Pretenure, Milgram told Jerome Bruner, a professor from Milgram's graduate program at Harvard University, "My hope is that the obedience experiments will take their place along with . . ." contributions by the biggest names in social psychology: "Sherif, Lewin and Asch" (as cited in Perry, 2012, p. 57). Whatever drove Milgram on, he anticipated enormous benefits for both scientific knowledge and himself. So what exactly did he find? What follows is a basic overview of his two baseline procedures and the counterintuitive results they produced.

N. Russell (✉)
University of Calgary, Calgary, AB, Canada



D. M. Allen, J. W. Howell (eds.), *Groupthink in Science*,
https://doi.org/10.1007/978-3-030-36822-7_20

## Milgram's Baseline Experiments

The first official baseline experiment involved an actor posing as a potential subject. He entered a laboratory and encountered an apparent scientist (another actor, hereafter called the experimenter). The ostensible subject was then introduced to a waiting naïve, and actual, subject. The experimenter then told both the actual and supposed subject that the experiment they volunteered to participate in was designed to investigate the effects of punishment on learning. One person was required to be the teacher and the other the learner. A rigged selection ensured that the actor/subject was always the learner, and the actual subject the teacher. The actual subject (now teacher) watched as the experimenter secured the learner to a chair and attached an electrode to his arm. The learner was informed that the subject, using a microphone from another room, would ask them questions regarding a word-pair exercise. The learner was able to electronically transmit his answers to the subject's questions.

The subject was then taken into an adjacent room and placed before the shock generator. This device had 30 switches aligned in 15-volt increments ranging from 15 to 450 volts. The experimenter instructed the subject to give the learner a shock for each incorrect answer proffered; and each incorrect answer warranted for the learner a shock one level higher than its predecessor. No shocks were actually administered.

Upon starting, the learner regularly provided incorrect answers and, as a result, acquiescent subjects quickly advanced up the switchboard. The experimenter responded to any signs of hesitancy by the subject with one or more of the following prods:

> Prod 1: Please continue, or, Please go on.
> Prod 2: The experiment requires that you continue.
> Prod 3: It is absolutely essential that you continue.
> Prod 4: You have no other choice, you must go on (Milgram, 1974, p. 21).

If the subject attempted to clarify the lines of responsibility, the experimenter asserted: "I'm responsible for anything that happens to him. Continue please" (p. 74). At the 300 and 315-volt shock switches, the learner banged on the wall and thereafter fell silent. This silence implied that the learner had at least been rendered unconscious. The experimenter then instructed the subject to treat all subsequent unanswered questions as incorrect and inflict a shock at the next level. The experiment was deemed complete upon the subject administering three successive 450-volt shocks. Sixty-five percent of subjects (26 out of 40) inflicted every shock.

After running this experiment, Milgram and his research team ran 23 variations. For example, for the fifth experiment, Milgram decided to run a second more radical "New" Baseline, where up until the 345-volt shock switch the subject could clearly hear the content of the learner's increasingly distressed reactions (eventual panicked screams) to being "shocked." The New Baseline condition also obtained a 65% completion rate, and thereafter became the model procedure that all subsequent slight variations were based on. During the final 24th "Relationship condition,"

subjects were encouraged to inflict increasingly intense shocks on an eventually screaming learner who was at least an acquaintance, often a friend, and occasionally a family member (see Russell, 2014b). To clarify, prior to the experiment's start learners were covertly informed of the study's actual purpose (Will your friend follow orders to hurt you?) and then instructed on how vociferously to respond to their friend's infliction of increasingly intense "shocks." This particularly unethical variation saw the completion rate plummet to 15 percent.

Data collection took 10 months and involved a total of 780 subjects (Perry, 2012, p. 1). The amount of data collected was enormous. Despite this, to date, nobody has managed to develop a "conclusive" theory capable of accounting for Milgram's findings (Miller, 2004, p. 233).

## Why Did Most Subjects Complete the New Baseline?

Despite the theoretic drought, it seems many factors, some of which I will describe below, are (perhaps cumulatively) likely to have contributed to most subjects' decision to complete the New Baseline experiment. The first such factor is termed "moral inversion" (Adams & Balfour, 1998, p. 20), which is where "something evil" (inflicting intense shocks on an innocent person) was converted by the experimenter into something "good" (advancing scientific knowledge on the effects of punishment on learning). The experimenter's higher "scientific" goals meant (apparently) the data *had* to be collected. As Milgram (1974, p.187) put it, the infliction of harm comes ". . . to be seen as noble in the light of some high ideological goal" where, by inflicting shocks, "science is served."

Another factor was the foot-in-the-door phenomenon, which is where persons are more likely to agree to a significant request if it is preceded by a comparatively insignificant request (Freedman & Fraser, 1966). For example, nearly every subject in the New Baseline inflicted the first six relatively light shocks (15–90 volts). However, in line with the foot-in-the-door phenomenon, doing so saw them comply with a small request which, unbeknownst to them, was about to be followed by some far more significant ones. The foot-in-door phenomenon is likely to have had two important consequences on subjects:

> (a) it engages subjects in committing precedent-setting acts . . . before they realize the "momentum" which the situation is capable of creating, and the "ugly direction" in which that momentum is driving them; and (b) it erects and reinforces the impression that quitting at any particular level of shock is unjustified (since consecutive shock levels differ only slightly and quantitatively). (Gilbert, 1981, p. 692)

Across many small 15-volt steps, most subjects inflicted increasingly intense and eventually dangerous "shocks."

Another likely influential factor over many subjects' decision to continue inflicting shocks was the undeniably coercive—even bullying—force of the experimenter's prods. The efficacious force of these prods was probably increased by the fact

that the experimenter—a scientist—was closely associated with Yale University—a highly credible and authoritative institution of knowledge.

The final influencing factor I will discuss here was the experimenter's offer to accept all responsibility for the subject's infliction of further shocks. This offer enabled a subject to displace responsibility for their shock-inflicting actions onto the experimenter and provided the subject with an important self-interested benefit: if the subject was (apparently) not responsible for their actions, then they were under no obligation to stop the experiment. Consequently, the subject could, at the learner's expense, avoid having to engage in the predictably awkward confrontation with the experimenter otherwise necessary to stop the experiment. That is, by accepting the experimenter's offer, the subject could continue flicking the switches and—(apparently) absolved of all moral and legal culpability—simply blame the experimenter for their actions (see Russell & Gregory, 2011).

The most many obedient subjects were willing to do to help the learner avoid the intensifying "shocks" was to covertly sabotage the experiment by verbally emphasizing to the learner the correct answers to the questions. Thus, these subjects were willing to sacrifice the (apparently) all-important scientific pursuit of knowledge in favor of their self-interested desire to avoid a confrontation with the experimenter.

It seems the cumulative effect of these forces—moral inversion, foot-in-the-door phenomenon, displacement of responsibility, and appealing to the subject's self-interested desires to avoid a confrontation—probably caused most subjects to fall in line with the experimenter's groupthink desires: inflicting further shocks was (apparently) essential. And once subjects totally committed to doing as they were told, their passing of this moral Rubicon saw some engage in some rather unusual behaviors. For example, on reaching the high end of the switchboard, some subjects started anticipating the learner's screams and then attempted to talk over them, thus actively trying to avoid having to hear (neutrialize) their pained appeals. These subjects—more concerned about alleviating their stress-related pain—did not want to know what they knew: that they had committed to hurting an innocent person but preferred to remain, as termed by Heffernan in a previous chapter, *willfully blind* to this reality.

At the earliest opportunity, Milgram attempted to and eventually succeeded in publishing the first official baseline experiment (1963). This publication, which mentioned the Holocaust in its first paragraph, garnered immediate media attention and with time became Milgram's "best-known result" (Miller, 1986, p. 9). Because he thought he had captured key elements of the Holocaust in the controlled laboratory setting, Milgram likely thought the wider academic community would heap praise on his research. But the first scholarly response, by Diana Baumrind (1964) in the prestigious *American Psychologist*, was a scathing ethical critique that also questioned the external validity of the untenured Milgram's experiment. Baumrind, for example, pointed out that unlike German perpetrators during the Holocaust, Milgram's typically concerned subjects clearly did not want to hurt their victim. Thus, she remained unconvinced by Milgram's generalizations towards the Holocaust. If Baumrind was right and no parallel to the Holocaust existed, then, as

she also notes, Milgram had no justification for having exposed his subjects to, as stated in his 1963 article, the following torturous experience:

> I observed a mature and initially poised businessman enter the laboratory smiling and confident. Within 20 minutes he was reduced to a twitching, stuttering wreck, who was rapidly approaching a point of nervous collapse. (Milgram, 1963, p. 377, as cited in Baumrind, 1964, p. 422)

If Baumrind was correct and harm was inflicted on innocent people for no reason at all, then the running of the Obedience studies were perhaps an example of groupthink captured in the laboratory setting. That is, through conformity and/or a desire for group harmony, somehow Milgram's many helpers—actors, research assistants, and technicians—all agreed to treat innocent people in an injurious and ultimately unethical manner. Other critics of Milgram, like Harré (1979, p. 106) for example, have alluded to this potentially groupthink connection:

> Milgram's assistants were quite prepared to subject the participants in the experiment to mental anguish, and in some cases considerable suffering, in obedience to Milgram. The most morally obnoxious feature of this outrageous experiment was, I believe, the failure of any of Milgram's assistants to protest against the treatment that they were meting out to the subjects.[1]

Milgram's research notes dated March 1962 showed he was aware of the ironic parallel between the subjects' *seemingly* harmful actions and his research team's *actually* harmful actions:

> Consider, for example, the fact --and it is a fact indeed, that while observing the experiment I ---and many others-- know that the naive subject is deeply distressed, and that…[it] is almost nerve shattering in some instances. Yet, we do not stop the experiment because of this […] *If we fail to intervene, although we know a man is being made upset; why separate these actions of ours from those of the subject, who feels he is causing discomfort to another. And can we not use our own motives and reactions as a clue to what is behind the actions of the subject. The question to ask then is: why do we feel justified in carrying through the experiment, and why is this any different from the justifications that the obedient subject's feel?* (Stanley Milgram Papers, Box 46, Folder 163.) [Italics added]

With his unrelenting ambition to develop a psychological (individual) theory capable of explaining why most subjects behaved in "a shockingly immoral way" (Milgram, 1964, p. 849), Milgram never further pursued this more sociological (group) and no doubt disconcerting observation.

To unravel why Milgram's research team agreed to inflict harm on innocent people, I would argue it is important to analyze the start-to-finish journey that led to Milgram's destination: his perplexing Baseline/New Baseline completion rates. This "behind the scenes" approach when viewing the actual running of the experimental program is, I believe, capable of revealing some of the more important sociological forces that encouraged Milgram and his research team's groupthink

[1] Unbeknown to Harré, one of Milgram's actors, Robert J. Tracy, refused to continue performing his acting duties. According to his son, Tracy "couldn't go through with it" and walked out (see Perry, 2012, p. 226).

decision to collect a full set of ethically questionable data. It is no coincidence, as we shall see, that these group forces coincide with those that likely affected the compliant subjects' decisions to participate.

## Group Forces Influencing the Research Team's Agreement to Inflict Intense Stress on Innocent People

When, after the first pilot study, Milgram decided to pursue the official Obedience research program, an obstacle likely to inhibit the realization of such ambitions became increasingly apparent. That is, because subjects during the first pilot experienced, as stated in his research proposal, "extreme tension" (as cited in Russell, 2014a, p. 412), there was a risk some of the specialists whose help he needed to collect the official data might deem the research program unethical and refuse to fulfill their essential roles.

Milgram's initial strategy to ensure that his research assistants, technicians, and actors all agreed to perform their roles was to encourage them to, as Fermaglich (2006, p. 89) put it, "view" the subjects' obedience "as an analogue of Nazi evil." Thus, much like he did with his subjects, Milgram morally converted "something evil" (imposing stress on the innocent subjects) into something "good" (generating scientific knowledge into better understanding perpetrator behavior during the Holocaust). The actor who most frequently played the role of the stress-inflicting experimenter, John Williams, for example, understood that despite his making "a man…upset," data collection was of "tremendous value," and thus the experiments "must be done" (as cited in Russell, 2014a, p. 416). Another example of moral inversion occurred when Milgram reassured his main research assistant, graduate student Alan Elms, that he did not need to worry about his "E[i]chman[n]-like" role of delivering a constant flow of subjects to the laboratory because they were all given "…a chance to resist the commands of a malevolent authority and assert their alliance with morality" (as cited in Blass, 2004, p. 99).

Although all helpers were encouraged to believe that they would be contributing to an important study, Milgram sensed that this in itself was not enough to secure everybody's long-term services. Thus, when necessary, he bolstered his moral inversion of bad into good by anticipating and then appealing to all his helpers' sometimes different self-interested desires. For example, Milgram offered actors Williams and James McDonough (the main "Learner") a generous hourly rate (which Milgram increased three times within eight months), along with the offer of a cash bonus to be paid out once all the data had been collected (Russell, 2014a, p. 416). Milgram also paid Elms an hourly rate for his services but also strengthened the attractiveness of role fulfillment by supporting the graduate students' emerging interest in the Obedience studies by publishing a journal article with him (see Elms & Milgram, 1966).

So, to promote involvement among all his helpers, Milgram basically applied what he suspected would prove to be the most successful individually tailored motivational formula—quid pro quo arrangements where benefits are provided in exchange for services rendered (Russell, 2014a, 416–417). Armed with typically similar justifications, it appears Milgram's helpers resolved the moral dilemma over whether or not to become involved in a potentially harmful study by becoming sufficiently convinced and/or opportunistically tempted into making their essential specialist contributions to data collection.

One might suspect that Milgram's helpers would have felt anxious about potentially harming innocent people, especially after weighing this risk up against the mere "scientific" and self-interested gains they hoped to obtain. This is especially so considering that during the official collection of data, at least two subjects were placed under such intense stress that they later complained that they thought they were going to have—or perhaps had—a heart attack (see Russell, 2009, pp. 104–105). However, alleviating such concerns was that as Milgram drew all his specialist helpers into role fulfillment, the issue of individual responsibility for harm infliction underwent a subtle yet powerful transformation. That is, after agreeing to perform their roles, all helpers unwittingly became links in an inherently stress-resolving and goal-directed assembly line-like bureaucratic process.

To clarify, before the official research program could proceed, Milgram had to design and then construct an inherently bureaucratic organizational process which would enable his research team to systematically and efficiently extracted data from 780 subjects. More specifically, "processing" involved training subjects, running the experiment, collecting data, and debriefing. For each subject, Milgram's research team had to complete all of these tasks within a pre-determined one-hour block so that the stage, so to speak, could be reset before the next subject's arrival at the top of the hour.

Intrinsic to all such bureaucratic processes is the division of labor (DOL)—where an organizational goal (in this case, collecting data) is subdivided into numerous tasks and then each of those tasks is allocated to a particular specialist functionary (Weber, 1976). For functionaries, however, this compartmentalization of tasks can cause a disjuncture between cause (for example, making partial contributions to Milgram's goal of collecting a full data set) and any negative effects generated by goal achievement (the infliction of intense stress on subjects). Among all functionary helpers—so-called cogs in the organizational machine—this disjuncture between cause and effect can stimulate what Russell and Gregory term "responsibility ambiguity" (2015, p. 136). Responsibility ambiguity is a metaphorical haziness, which renders debatable which functionary helper is most responsible for any harm inflicted by the wider organizational process. Importantly, responsibility ambiguity makes it difficult for arbiters to later determine who should be held to account for such harmful outcomes. This haziness can render some functionary helpers genuinely unaware of their personal responsibility. However, this haziness can also enable others to opportunistically escape shouldering responsibility because they suspect that their harmful contributions will be rewarded in the short-term and, due to the availability of plausible deniability ("I didn't know!"), never punished in the

long-term. Therefore, it could be argued that the bureaucratic process structurally provided all of Milgram's helpers with the "fog" of responsibility ambiguity (Russell & Gregory, 2015).

Perhaps the most common source of responsibility ambiguity among functionary helpers working across an organizational chain is the option to displace or "pass the buck" of responsibility for their harmful contributions elsewhere (Russell & Gregory, 2015). For example, had a subject been seriously injured during data collection, Williams the stress-inflicting experimenter could, if he so chose, blame Milgram for his actions: Williams was only following his employer's instructions. Milgram, the principal investigator, was only undertaking the kind of groundbreaking research that prestigious universities like Yale pressured non-tenured faculty into pursing: he too was only doing his job. Perhaps the funders of the research—the National Science Foundation (NSF)—or the chair of Yale's Department of Psychology, Claude E. Buxton (Milgram's boss), were most responsible: they ultimately allowed, desired, and legitimized Milgram's research. The NSF and Buxton, however, did not directly hurt anyone and they certainly never condoned Milgram's pursuit of the particularly unethical Relationship condition. Perhaps, in the end the reified ideological pursuit of "scientific knowledge" was mostly to blame. The point is, as soon as a bureaucratic process forms, it suddenly becomes possible for all functionary helpers to blame someone or something else for their contributions to a harmful outcome. And because "others" were involved, it seems all sensed they could probably make their individual contributions with probable impunity. And on all realizing this, every helper thereafter only needed to concern themselves with reaping the personal benefits on offer for making their specialist contributions. This may help explain why Milgram's helpers risked partaking in such a potentially dangerous experiment.

Another subtle yet powerful effect the DOL can have on functionary helpers is termed bureaucratic momentum (Russell & Gregory, 2015). Bureaucratic momentum has usually taken hold when functionaries experience pressure to perform their specialist roles by preceding and sometimes succeeding functionary links across an organizational chain. This coercive force appears to be generated by the cumulative momentum of the many simultaneously moving functionary "cogs" bearing down and exerting pressure on one another. Functionary links often experience this coercive force to fulfill their roles in the form of peer pressure: "to get along" one must "go along." For example, in fear of causing a bottleneck or delay in organizational goal achievement, employees on a factory assembly line typically feel pressure to quickly fulfill their specialist roles. A single uncooperative functionary can—say because of moral reservations—resist such pressure; although doing so is rare because they must sacrifice whatever self-interested benefits they might otherwise have received for performing their specialist role. Also, this kind of resistance deprives other (potentially angry) functionaries from obtaining whatever benefits they anticipated receiving for organizational goal achievement. It is less stressful on everybody involved if all give in to the momentum of role fulfillment and just do their bit for goal achievement.

Bureaucratic momentum is likely to have had an influential effect during the Obedience studies. For example, to please his funders at the NSF, Milgram likely felt pressure to collect—despite any emerging ethical reservations—a full set of data. Doing so, however, required the long-term retention of the experimenter's acting services. In return for being retained over a long period of time, the experimenter—despite any emerging ethical reservations—likely felt contractually obliged to continue placing subjects under enormous stress. And, of course, it could be argued that the experimenter's seemingly unrelenting prods, like it having been "absolutely essential" the subject "continue" inflicting more shocks, saw—despite any emerging ethical reservations—the transfer of bureaucratic momentum to the last functionary link in the Obedience study's data collecting organizational chain.

The final group force I'll mention here likely to have influenced Milgram's research team was (again) the foot-in-the-door phenomenon. For example, it could be argued that after Milgram's research team agreed to undertake the first official and, relatively speaking, benign (first) Baseline condition (where the learner banged a few times on the wall), the more amenable (or perhaps desensitized) the team became to undertaking the fifth more radical New Baseline experiment (where an increasingly hysterical learner suddenly went silent). With the entire research team having agreed to undertake the more radical New Baseline, the more amenable they became to undertaking the most radical 24th and final Relationship condition where, as mentioned, subjects were pushed to inflict severe "shocks" on someone who was at least an acquaintance, often a friend, and sometimes a family member. The point being, it is unlikely Milgram's helpers would have had the nerve to run the Relationship condition at the start of the data collection process. The slippery slope of the foot-in-the-door phenomenon—small and barely perceivable steps in an increasingly radicalized direction—likely had a powerful influence on those working within the Obedience study's data-collecting bureaucracy.

In summary, much like with the obedient subjects, the forces of moral inversion, receiving self-interested benefits, displacement of responsibility, bureaucratic momentum, and the foot-in-the-door phenomenon all (perhaps cumulatively) likely exerted an influence on the research team's groupthink decision to collect a full set of ethically questionable data.

## Prioritization of Milgram's Self-Interests over the Scientific Pursuit of Knowledge

It seems the reason Milgram decided to run the experimental program was because he believed the benefits—greater knowledge into mankind's destructive tendency to obey—outweighed all the costs. As he said in the draft notes of his 1974 book:

> Under what conditions does one ask about destructive obedience? Perhaps under the same conditions that a medical researcher asks about cancer or polio; because it is a threat to human welfare and has shown itself a scourage [sic] to humanity. (As cited in Russell, 2009, p. 104).

But, when Milgram decided to pursue his research program, it seemed the only people faced with paying any "costs" would be his obedient subjects (whom, as far as he was concerned, only got what they deserved for, as mentioned, failing to "assert their alliance with morality"). Again, he, on the other hand, could only envision personally benefiting from running the official experiments. But after the publication of Baumrind's (1964) critique, this all suddenly changed.

Baumrind's critique rather suddenly threatened to label his research unethically abusive and perhaps even held the potential to destroy his fledgling academic career. With *his* personal self-interests suddenly on the line, Milgram realized *he* might have to pay a high price for his earlier decision to proceed with the study. With his back against the wall—and much like those subjects who attempted to sabotage his experiments—Milgram *also* started prioritizing his self-interests over and above the so-called importance of generating scientific knowledge. That is, post-Baumrind, Milgram set about protecting his personal interests by compromising the accuracy of the knowledge he had collected—what he did and found during data collection—by massaging the truth, omitting certain facts, and even telling complete lies. Thus, like many examples of groupthink, the emergence of certain negative outcomes was followed by a carefully calculated cover up.

For example, despite encountering subjects complaining about their hearts, in his response to Baumrind (and repeatedly thereafter) a perhaps *willfully blind* Milgram (1964, p. 849) described his subjects' stress as mere "momentary excitement," a sudden change in tone that Patten (1977, p. 356) observed to be "a most astonishing about-face." In his book, Milgram noted that before each trial subjects had to sign "a general release form, which stated: 'In participating in this experimental research of my own free will, I release Yale University and its employees from any legal claims arising from my participation'" (1974, p. 64). But what he failed to disclose was, as stated in his personal notes, "The release, of course, was not used for experimental purposes, but to protect us against legal claims" (as cited in Russell, 2014a, p. 418). If Milgram honestly believed his experiments only caused "momentary excitement," why did he need legal protection?

Another omission was that although before Baumrind's critique Milgram promised to publish the Relationship condition's results, after her critique he mysteriously never mentioned the variation again (Russell, 2014b). Of course, if Baumrind's critique of the relatively benign first Baseline could, as Milgram clearly sensed, threaten the reputation of his research, one can only imagine the ethical firestorm she would have unleashed on him had he published a variation where some subjects were pushed into inflicting harmful "shocks" on a relative. And in terms of outright lies, Milgram counter-critiqued Baumrind for confusing "the unanticipated outcome of an experiment with its basic procedure," then elaborating that "the extreme tension induced in some subjects was unexpected" (Milgram, 1964, p. 848). Milgram said this despite him having earlier undertaken numerous pilot studies where, as mentioned, some subjects experienced what he termed in his research proposal "extreme tension".

It can therefore be argued that Milgram's self-interests—protecting his name, career, and the ethical reputation of his world-famous experiments—ended up being prioritized over (and thus ultimately corrupted) his espoused purest beliefs surrounding the so-called scientific pursuit of knowledge. Cementing this chapter's focus on the overlap between individual and group behavior during the Obedience experiments, at some level Milgram self-reflexively sensed a connection between his obedient subjects' self-centered decisions to prioritize their personal interests over the well-being of the learner and him prioritizing his self-interests over the subjects' well-being:

> Moreover, considered as a personal motive of the author --the possible benefits that might redound to humanity --withered to insignificance along side [*sic*] the strident demands of intellectual curiosity. When an investigator keeps his eyes open throughout a [scientific] study, he learns things about himself as well as about his subjects, and the observations do not always flatter. (As cited in Russell, 2009, p. 186)

## Conclusion

Milgram naturally viewed himself as a detached, objective, and scientific observer of destructive social behavior. That is, he set up an experiment but perceived himself to be independent of the results it produced. He, however, failed to sense his own highly involved non-scientific role in the social engineering of those results. Two particular factors he remained oblivious of were, first, the subtle power inherent within the data-extracting bureaucratic process he constructed (and the necessary role it played in helping generate his surprising results— a key structural force that likely explains much of the ironic overlap in group and individual behavior). The largely invisible role of bureaucratic organization no doubt plays a key role in helping socially engineer many other "real life" examples of groupthink behavior—particularly because of its ability to promote, among all functionary links across the chain of command, feelings of responsibility ambiguity. Second, Milgram was largely unaware of the important role that his and his research team's self-interests played in both helping generate the surprising results and corrupting their scientific pursuit for new knowledge. This last point may have implications that extend beyond Milgram's laboratory walls. For example, what role did the pushes and pulls of bureaucratic organization and personal self-interest play in stifling dissent among some of the scientists working on the Manhattan Project? Finally, I am confident that Milgram's dissectible research—somewhat uniquely captured in the (semi)controlled social science laboratory—is likely to provide scholars with great insights into the inner workings of other more contemporary examples of highly destructive and seemingly unstoppable groupthink behavior, like for example, climate catastrophe.

## References


Adams, G. B., & Balfour, D. L. (1998). *Unmasking bureaucratic evil*. Thousand Oaks, CA: Sage Publications.

Baumrind, D. (1964). Some thoughts on ethics of research: After reading Milgram's 'behavioral study of obedience'. *American Psychologist, 19*(6), 421–423.

Blass, T. (2004). *The man who shocked the world: The life and legacy of Stanley Milgram*. New York, NY: Basic Books.

Elms, A. C., & Milgram, S. (1966). Personality characteristics associated with obedience and defiance towards authoritative command. *Journal of Experimental Research in Personality, 1*(4), 282–289.

Fermaglich, K. (2006). *American dreams and Nazi nightmares: Early Holocaust consciousness and liberal America, 1957–1965*. Waltham, MA: BrandeisUniversity Press.

Freedman, J. L., & Fraser, C. C. (1966). Compliance without pressure: The foot-in-the door technique. *Journal of Personality and Social Psychology, 4*(2), 195–202.

Gilbert, S. J. (1981). Another look at the Milgram obedience studies: The role of the gradated series of shocks. *Personality and Social Psychology Bulletin, 7*(4), 690–695.

Harré, R. (1979). *Social being: A theory for social psychology*. Oxford, UK: Basil Blackwell.

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology, 67*(4), 371–378.

Milgram, S. (1964). Issues in the study of obedience: A reply to Baumrind. *American Psychologist, 19*(11), 848–852.

Milgram, S. (1974). *Obedience to authority: An experimental view*. New York, NY: Harper and Row.

Miller, A. G. (1986). *The obedience experiments: A case study of controversy in social science*. New York, NY: Praeger.

Miller, A. G. (2004). What can the Milgram obedience experiments tell us about the Holocaust? Generalizing from the social psychology laboratory. In A. G. Miller (Ed.), *The social psychology of good and evil* (pp. 193–237). New York, NY: Guilford Press.

Patten, S. C. (1977). The case that Milgram makes. *The Philosophical Review, 86*(3), 350–364.

Perry, G. (2012). *Beyond the shock machine: The untold story of the Milgram obedience experiments*. Melbourne: Scribe.

Russell, N. (2014a). The emergence of Milgram's bureaucratic machine. *Journal of Social Issues, 70*(3), 409–423.

Russell, N. (2014b). Stanley Milgram's obedience to authority "relationship" condition: Some methodological and theoretical implications. *Social Sciences, 3*(2), 194–214.

Russell, N., & Gregory, R. (2011). Spinning an organizational "web of obligation"? Moral choice in Stanley Milgram's "obedience" experiments. *The American Review of Public Administration, 41*(5), 495–518.

Russell, N., & Gregory, R. (2015). The Milgram-holocaust linkage: Challenging the present consensus. *State Crime Journal, 4*(2), 128–153.

Russell, N. J. C. (2009). Stanley Milgram's obedience to authority experiments: Towards an understanding of their relevance in explaining aspects of the Nazi Holocaust. PhD thesis: Victoria University of Wellington.

Weber, M. (1976). *The Protestant ethic and the Spirit of capitalism*. Translated by T. Parsons with an introduction by A. Giddens. London: Allen & Unwin.

# Chapter 21
# The Physician's Dilemma: Healthcare and Bureaucracy in the Modern World


**J. Kim Penberthy and David R. Penberthy**

## Introduction

Physicians are educated in the science of medicine and healing of patients and have traditionally been less oriented toward the business aspects of healthcare. Perfectionism, hard work, and sacrifices by physicians have helped to advance healthcare in the United States and around the globe over the past several decades. Along with the exponential scientific growth, expanded healthcare options, and growing complexities that arise, modern medicine has become increasingly regulated. This has resulted in expansion of mandatory requirements and an explosion in the growth of administrators and bureaucrats who, by the very nature of their training, focus on the "bottom line" of outcomes and costs of healthcare.

Competing incentives between physicians' way of approaching medicine and the bureaucrats' approach to healthcare has created tension for both. Physicians are feeling increasingly disenfranchised within the practice of medicine due to a myriad of factors. It is no wonder then, that physician burnout has risen over time. This burnout has been associated with distressed and disruptive physician behaviors and negative impacts on the healthcare environment and patient care.

We propose that "burnout" and related distressed behaviors of physicians are best conceptualized as a symptom of the overall dysfunction within the healthcare system. These inter- and intrapersonal mechanisms have resulted in a form of physician groupthink characterized by indignant frustration, helplessness, and inaction. We propose that the key to addressing physician burnout is larger than merely teaching physicians mindfulness strategies or improved coping skills. The goal

J. K. Penberthy (✉)
University of Virginia School of Medicine, Charlottesville, VA, USA
e-mail: Jkp2n@virginia.edu

D. R. Penberthy
Virginia Radiation Oncology Associates, Richmond, VA, USA



D. M. Allen, J. W. Howell (eds.), *Groupthink in Science*,
https://doi.org/10.1007/978-3-030-36822-7_21

cannot be to simply train physicians how to endure an increasingly burdensome and nonsensical healthcare system. We propose that physicians must instead reshape their role in healthcare, become more proactive, engage in leadership, and advocate for their profession, including pursuing common-sense approaches to treatment and increased autonomy over patient care. These behavioral and advocacy changes may renew physicians' energy and decrease burnout within the field of medicine and may even lead to reduced administrative burdens and costs, thus effecting lasting positive change in healthcare.

## The Complex Character of Physicians

People who choose careers in medicine have traditionally demonstrated personality traits and aptitudes that include high intelligence, compassion, inquisitiveness, and sensitivity to others. They are also typically competitive, driven, independent, and perfectionistic (Lemaire & Wallace, 2014). The impact of such a well-intended group of high-achieving and hardworking individuals in the field of medicine has been a blessing for patients and the field of healthcare and potentially one cause of the development of a "groupthink" in physicians which has contributed to their current increased rates of dissatisfaction and burnout.

Individuals attracted to a career in medicine are motivated to help others in practical and significant ways. Most physicians understand that a career in medicine means a lifetime of service to their community and they enter this societal contract willingly. To become a physician, one must have the intellectual curiosity, capacity, endurance, and perseverance to get into and successfully complete medical school and residency programs. Simultaneously, due to the nature and intensity of the work, physicians are often more socially isolated and emotionally disconnected than individuals in other professional fields (Lipsenthal, 2005). Pre-medical collegiates are studious and independent, perhaps spending more time in the library than socializing with others. They are individuals who are achievement oriented, self-motivated, deeply engaged in academics, and invested in individual academic success (Eley, Leun, Hong, Cloninger, & Cloninger, 2016). Medical school and residency continue the indoctrination of this self-selected group.

The humanistic component of clinical competence, such as empathy and other interpersonal skills, can be eroded in medical students who are vulnerable to the rigors of medical school (Hojat, Spandorfer, & Mangione, 2013). These institutions exacerbate the natural tendencies of the individuals by making 80-hour work weeks normative and underscoring individual responsibility and achievement throughout the group. While their non-physician peers are enjoying active social lives, more pay, less demanding jobs, or starting families, young physicians in training are working in isolation and under extreme stress, collecting increasingly large debt, and making little income, with less autonomy and free time.

Most training programs are also intensely hierarchical with attending physicians setting the pace and expectations for more junior physicians in training. In our experiences, challenging senior physicians in medicine is not typically considered an option and is certainly not encouraged. If young doctors-to-be are afraid to discuss their daily challenges, learning challenges or mistakes, and fears, the isolation only becomes more extreme.

Conventional medical training, although making incremental improvements, is still highly dysfunctional. The training process exaggerates the individual's already described traits, often at the expense of innovation and social connection. It seems a basic tenet of medical training that if trainees are overloaded with responsibility and information, they will rise to the occasion, and most do. Physicians in training work long hours, get paid barely enough to survive (especially given their financial debts), and still must see ever more patients and meet other research and service demands. Too often, this leads to increased isolation or emotional distance. This tendency for emotional dissociation may be further developed in the anatomy lab, emergency room, and other places and situations where emotions are not so helpful. Physicians may learn to shut off emotions, keeping the "scientist" mindset.

These are all useful mechanisms, allowing physicians to do difficult but necessary tasks. However, such strategies often lead to additional isolation and a feeling of being "in it alone." Many physicians exhibit compulsive traits, especially what has been called the "compulsive triad" of self-doubt, guilt, and an exaggerated sense of self-importance (Spickard, Gabbe, & Christensen, 2002). Self-doubt often results from having excessively high personal standards, common in many physicians, that are often so high that the standards are difficult, or impossible, to achieve. Given these high self-expectations, such physicians often impose equally high standards on others and react strongly if colleagues or staff fail to meet them. There is some evidence that physician training and work is indeed so stressful that many physicians may meet criteria for a type of chronic stress disorder (West, Shanafelt, & Kolars, 2011). The key symptoms are intrusive thoughts, avoidance behaviors, and hyperarousal. Learned helplessness has also been hypothesized to be a factor in distressed physicians, who despite their best efforts, cannot seem to stay ahead of the workload.

Despite these stressors and pressures, the majority of physicians successfully finish their training and enter into their careers with a focus on patient care and service. They understand the commitment that medicine takes and willingly enter into this world. However, they may be forever changed by their experiences in training, and certainly many of them have learned behaviors to help them survive – including a skewed expectation of an intense workload, emotional dissociation, and little to no expectation of improvement in their situation. Many of them are chronically stressed, financially strapped, and may feel helpless (Thomas-Dyrbye & Shanafelt, 2006).

Healthcare has benefited from the hard work and dedication of such physicians. Quality medical care and prevention has blossomed, with increased access and improved treatments across the globe (Berwick, Calkins, McCannon, & Hackbarth, 2006).

Physicians work long and hard to practice a challenging profession that is also incredibly satisfying and rewarding to most of them. However, with increased success and expansion of healthcare have come increased legislation and bureaucracy, with increased administrative and management work, and the need for individuals to navigate these complexities. Thus, enter the healthcare administrator, manager, and other bureaucrats.

## Increasing Role of Bureaucracy in Medicine

According the Centers for Medicare and Medicaid Services (CMS; National Healthcare Expenditures Fact Sheet, 2018), the U.S. National Health Expenditures grew to $3.3 trillion in 2016, or $10,348 per person, and accounted for 17.9% of U.S. gross domestic product (GDP). CMS projects an annual growth of 5.5%, meaning about one-fifth of U.S. GDP will be spent on healthcare by 2025 with overall GDP projection of a total economy of $25 trillion, which means about $5 trillion will be spent on the healthcare system that year. Our government along with health insurance companies, hospital systems, and other agencies has created bureaucracies to manage and direct all of this money for healthcare.

Growing numbers of healthcare regulations lead to an increased need for management to ensure compliance with these well-intentioned rules. Such work necessitates time and attention most physicians do not have due to their patient care obligations and work hours. Additionally, many physicians may not have the business knowledge and management skills to be competitive or successful in the field of healthcare administration. Thus, the healthcare system in the United States has witnessed a staggering rise in the number of non-physician administrators and managers over the past decades. The numbers from 1975 to 2010 are dramatic, to say the least, with a 3200% increase in the number of administrators compared to 150% increase in the number of physicians over that time period (Cantlupe, 2017). For perspective, the increase in physician numbers roughly kept up with population growth over this 35-year period.

Supporters say the growing number of administrators is needed to keep pace with the drastic changes in healthcare delivery over the past decades, particularly change driven by technology and by ever-more-complex regulations. To cite just a few industry-disrupting regulations, consider the Prospective Payment System of 1983, the Health Insurance Portability & Accountability Act of 1996, the Health Information Technology for Economic and Clinical Act of 2009, and The Patient Protection and Affordable Care Act of 2010. Critics say the army of administrators does little to relieve the documentation burden on physicians, while creating layers of high-salaried bureaucratic bloat in healthcare organizations.

Physicians now spend roughly two-thirds of their professional time on paperwork – mostly filling out the never-ending fields that are part of Electronic Medical Records requirements – rather than attending to patients (Sinsky et al., 2016). Remember also, that physicians do not get reimbursed for completing paperwork.

This means patients are essentially spending three times more than they should have to for their doctors' time. Simply halving doctors' paperwork could halve physicians' costs because they would have more time for productive, patient-centered work. Hospital costs are highest in the countries that have the highest administration costs (Bouchard, 2014). Research supports the fact that increased numbers of administrators is associated with increased cost of healthcare but not improved outcomes (Woolhandler, Campbell, & Himmelstein, 2003).

Hospital administrators are vital to ensuring that medical facilities run efficiently and deliver quality care, which appears to be in alignment with the goals of physicians. However, despite the appearance of alignment, differing incentives – both positive and negative – have created a disconnect between physicians and administrators. Hospital management teams are well versed in metrics including market share, revenue, and costs. They are aware when the hospital is operating with a surplus or not and are motivated to increase patient numbers and overall budget surplus.

Physicians place high value on quality patient care and will work hard for their patients. However, due to a myriad of factors including a loss of autonomy regarding patient care, requirements to complete large amounts of seemingly irrelevant or unnecessary paperwork, longer work hours, decreasing financial reimbursement, increasing threats of lawsuits, and frequent understaffing or lack of qualified and experienced ancillary healthcare workers, physicians may feel increasingly disenfranchised with the healthcare system and frustrated with their profession (Lathrop, 2017). These competing incentives between physicians' way of approaching medicine and the bureaucrats' approach to healthcare has created tension for both and has only added to the crisis in healthcare (Levine & Gustave, 2013).

## Physician Distress and Burnout

What happens when people with the personalities we described – perfectionistic, high-achieving, and independent – are put under additional stress, especially when they are given immense responsibility and very little authority? This describes what has happened to physicians in modern medicine today.

Physicians are confronted with fewer resources, increasing government regulations, greater patient outcome expectations, and rising student debt (Privitera, Rosenstein, Plessow, & LoCastro, 2015). There is also more pressure to practice in specific ways, such as adhering to guidelines and pathways that limit physician autonomy, and ongoing threats of lawsuits and liability. Many physicians express dissatisfaction with the decreasing amount of time allocated to each patient and consider their workload "too heavy" (Rosenstein, 2017). Satisfaction with work-life balance has significantly declined in physicians (Shanafelt et al., 2015).

In a 2014 American Medical Association national survey, 54% of practicing physicians met criteria for burnout (Shanafelt et al., 2015). It should be noted that the issue of distress affects nearly every group of physicians ranging from interns

(Dyrbye et al., 2014; Rosen, Gimotty, Shea, & Bellini, 2006; Shanafelt, Sloan, & Habermann, 2003) to department chairs (Gabbe, Melville, Mandel, & Walker, 2002). These ever-growing strains, coupled with a competitive and demanding work environment, have led to numerous negative psychological consequences including burnout and, in some cases, suicide (Schernhammer & Colditz, 2004). Overall, physician burnout has been associated with distressed and disruptive physician behaviors as well as negative impacts on the healthcare environment and patient care (Dewa, Loong, Bonato, & Trojanowski, 2017; Rosenstein, 2015).

This increased awareness of burned out and distressed physicians has not necessarily led to increased organized efforts to address the dysfunctions in the healthcare system but instead led to a cottage industry of programs and strategies to help educate or train the physician regarding tools to help them improve coping skills or teach them to be more mindful. This approach assumes or implies that the distress, dysfunction, or burnout is the fault of the physician and their lack of abilities or perhaps their nefarious motivations.

The poor behavior of the physician is often attributed solely to the physician's lack of abilities and several programs around the country have been developed to help promote increased emotional intelligence, effective coping, and mindfulness skills in physicians. While these may be helpful skills for physicians as well as any other human being, we argue that they do not address a crucial component of physician burnout, which is related to systems issues in the modern U.S. healthcare system. These issues include unnecessary bureaucratic and paperwork burdens, ever changing and uncertain health insurance regulations, and increasing lack of autonomy of physicians to perform the advanced diagnostics, procedures, and treatments for which they trained long and hard. We posit that too often these very real underlying issues are ignored or minimized in lieu of labeling the physician as distressed, disruptive, or burned-out and advocating for education of the individual instead of reformation of the system.

The situation that physicians find themselves in can be conceptualized as a form of groupthink on the behalf of the physicians who unnecessarily accept the label of "disruptive" or "distressed" physician and continue to complain about their burdens while making no overtures to address the real underlying issues. In fact, they may not even address the presenting issue of burnout – a 2012 study revealed that 78.3% of the distressed physicians surveyed had not previously thought about seeking professional help for distress or burnout (Fridner et al., 2012)!

## Challenges

Why are not physicians rising to the challenge and helping to change the current dysfunctional healthcare system? What is it about their groupthink that keeps them in such a dilemma? In the current healthcare system, most healthcare is delivered in a reactive way. Patients present with medical issues, sickness, and disease, and

physicians manage their condition. In the real world, with power in the healthcare system increasingly concentrated in the administrator level, physicians have less control over day-to-day clinic and hospital operations, policy, and patient care. Control over these central issues is held by administrators and managers, the vast majority of whom have no medical training (which in and of itself, may be infuriating to physicians). These administrators implement the business models in which they were trained and the hospital or clinic is run as a business enterprise with compliance to associated regulations and policies as a focus.

Today's physicians enter into this rigid business-focused system with their own entrenched groupthink of patient-focused care along with learned expectations of perfection, dedication, and perhaps a heavy dose of learned helplessness and self-doubt, as previously described. We propose that this combination is a part of the overall problem in modern healthcare. Additional challenges for physicians include staggering student loan debts that must be paid off and thus, their focus is on maintaining employment to stay financially viable. Many are also trying to start or keep families after years of isolating training.

Some physicians may forgo the bureaucracy of insurance and provide concierge medicine, only to face their own ethical dilemma of violating their own values by "abandoning" underprivileged populations. The personality traits and learned habits of physicians may render them more likely to honor the perceived hierarchy of authority in the hospital (as they were taught to do in training) and to try and solve issues on their own or outside of the system. This may help explain why an increasing number of physicians report feeling disenfranchised with the day-to-day work of their medical practice, yet seem to do little to directly or effectively improve the situation (Dewa et al., 2017). Passivity of physicians seems to have only lead to more bureaucracy, and those physicians who do speak up may be labeled "disruptive" or "burned out" by their administration (Reynolds, 2012). Even when well-intentioned, a physician who expresses dissatisfaction with the current state of affairs could be labeled a disruptive physician, with potential significant consequences (e.g., peer review processes, costly training programs, loss of clinical privileges), which can affect their ability to work. There is an increased sense of learned helplessness regarding the physician's ability to change "the system." However, that is exactly what today's physicians must do.

## Solutions

We propose that physicians' current way of thinking is not productive and is potentially harming physicians and the healthcare system. Effective solutions must include physicians working collectively to overcome the collective thought that they are powerless in the current system and asserting more control in the healthcare arena. We realize that the current healthcare system is extremely complex and ever evolving and that there is no one "magic bullet" that will solve the problem.

We propose that solving the challenges facing American healthcare will require a distinctly different type of relationship between physicians and administrators than currently exists in most health systems. This may involve helping physicians learn new skills to enhance the communication and emotional intelligence skills that they already possess. Many of these programs are currently available for physicians but are often only offered when the physician is already in trouble or having problems. We propose that offering leadership courses proactively or in medical school could help arm physicians with skills to better overcome negative groupthink tendencies and enhance wellbeing. Well-being should be considered more than simply the absence of distress.

Programs teaching mindfulness, effective communication skills, and stress reduction techniques may be key in helping to establish a resilience and effective group of physicians. Physician engagement in mindful communication programs has been associated with both short- and long-term physician well-being and positive attitudes associated with patient-centered care (Krasner et al., 2009). Mindfulness-based programs for physicians have demonstrated reduced burnout levels that may ultimately lead to a reduction in groupthink characteristics of over-achievement, guilt, and avoidance (Goodman & Schorling, 2012). Increasingly, medical schools are including mindfulness education and beginning to explore the impact on physicians (Dobin & Hutchinson, 2013). Findings indicate that more research is needed and that targeted interventions may be needed to impact specific maladaptive groupthink characteristics of physicians (Daya & Hearn, 2018). These strategies alone, however, are not enough to help physicians speak up, participate, and make the dramatic and lasting changes needed in today's healthcare system. In fact, a singular focus on improving physicians' coping and interpersonal skills risks laying the sole blame and responsibility on physicians, which is not the case in such a complex system.

Positive and lasting improvements in healthcare will also entail proactive participation in administration by physicians, including physician leadership at all levels. This will necessarily entail a shift in the groupthink of physicians currently in the workforce, and thus involvement of students and early career physicians is important. One immediate strategy that may be employed is to formalize processes and structures to tap the ingenuity, innovation, and knowledge of practicing physicians. As health systems focus increasingly on maximizing value, physicians are dramatically underutilized assets. Health systems and hospitals build broad-based committees and coalitions, but there are often no physicians on them or those that are included are part-time administrators with little current clinical experience. Practicing physicians may have access to real-time knowledge and insight into problems and solutions in healthcare delivery that administrators lack. In order to implement this, there will need to be time allotted for physicians to participate in committees, and physicians will need to commit to attendance and participation. Research has demonstrated that physician involvement in strategic decision making and investments in operational capabilities are associated with improved hospital performance (Goldstein & Ward, 2004). This type of empowerment of physicians can lead to genuine insights that enable improved care and cost savings.

Another strategy is to educate physicians about the financing of healthcare in order to allow them more knowledge, power, and insight into this area of medicine. Most physicians complete their training with little or no knowledge about the financing or organization of healthcare. Nowhere in their premedical education, medical school, residency, or fellowship do they get a comprehensive education on healthcare policy, administration, finance, or organizational behavior (Mou, Sharma, Sethi, & Merryman, 2011). This can lead to mistrust and suspicion on the part of physicians especially if they have bought in to the groupthink of competitiveness, helplessness, and doubt. Spanning this gulf of knowledge can go a long way to help rebuild understanding and trust.

Physicians are increasingly interested in understanding the finance and business of medicine but only if they are actively involved in decision making and feel that their voices are heard and honored (Jain & Miller, 2012). There has been a substantial growth in physicians obtaining their Masters of Business Administration (MBA) over the past decades (Gorenstein, 2017). However, research demonstrates that after completing their education, a majority of physician-MBAs divert their primary professional focus away from clinical activity (Ljuboja et al., 2016). Progressive health administrators must invest in preparing physicians to understand how healthcare is paid for and how payment informs the structure of care delivery. Absent this understanding, there will always be a layer of mistrust and confusion that gets in the way of true constructive dialogue and engagement about how to solve problems of healthcare delivery.

The complementary idea to providing business and healthcare administration education to physicians is to teach administrators about clinical medicine. This does not necessitate that administrators obtain an M.D. or D.O., but that they are schooled enough in clinical medicine so that they can better understand the complexities of clinical care and better speak to the issues in a common language as their physician counterparts. At a minimum, administrators could better understand how care is organized and delivered on the front lines through intensive clinical shadowing that can help create mutual understanding and perhaps engender respect. Just such a thing was initiated at Mission Health in Asheville, North Carolina, where they created an "immersion day" for their board members, journalists, legislators, and regulators to experience a day at the hospital and clinics in scrubs, behind the scenes, immersed in the nuances of care delivery (Bock & Paulus, 2016). The organizers of the Mission Health project included in their article a statement from a non-physician board member who stated: "I learned more about hospitals and health care from my 10 immersion hours than 6 years sitting on our board" (Bock & Paulus, 2016, p. 1202).

What if those involved in the financing and administration of healthcare delivery came to physicians from a place of increased knowledge and respect? This would potentially go a long way to help physicians overcome their groupthink tendencies and more effectively engage to create real and lasting change. We suspect that solutions that administrators and physicians design together would then be more patient-centered and more likely to deliver value than those either side would develop alone.

Obviously, there are scores of other components of the healthcare system that could be addressed to help reduce physician and patient burden, streamline the system, and improve patient delivery and care. These include a laundry list of things to change: reduce administrative burdens of physicians, make electronic medical records more useful with less unnecessary documentation, allow increased time for adequate patient care, streamline health insurance, address liability and legal issues, increase price transparency, and reduce unnecessary regulations, to name a few.

We propose that addressing the foundational issues of physician groupthink in order to help facilitate physician wellness and improve communication between physicians and administrators are the first necessary steps to help pave the way to solving these other issues. Physicians must take back the leadership roles in medicine and healthcare and do to so, they must lay aside the groupthink characteristics that have landed them in their current dilemma. Physicians alone cannot solve healthcare's biggest problems without collaborating with talented, dedicated, and multidisciplinary administrators. Nor can these administrators solve the same problems without the robust and thoughtful engagement of physicians. Some of what we are proposing is already happening, and one big effort in particular is worth noting. Industrial heavyweights Jeff Bezos, Warren Buffett, and Jamie Dimon, with over one million employees within organizations they lead, in January 2018 announced the formation of a healthcare initiative. On June 20, 2018, they announced Atul Gawande, M.D., as the CEO of this as-yet unnamed healthcare initiative. This intentional collaboration between businessmen and practicing physicians is exactly what we believe is necessary to improve today's healthcare system.

## References

Berwick, D. M., Calkins, D. R., McCannon, C. J., & Hackbarth, A. D. (2006). The 100000 lives campaign: Setting a goal and a deadline for improving health care quality. *Journal of the American Medical Association, 295*(3), 324–327.

Bock, R. W., & Paulus, R. A. (2016). Immersion day – Transforming governance and policy by putting on scrubs. *New England Journal of Medicine, 374*, 1201–1203.

Bouchard, S. (2014). Hospital budgets consumed by bureaucracy. *Healthcare Finance News.* www.healthcarefinancenews.com/new/hospital-budgets-consumed-bureaucracy

Cantlupe, J. (2017). The rise (and rise) of the healthcare administrator. *Athena Insight.* www.athenahelht.com/insight/expert-forum-rise-and-rise-healthcare-administrators

Center for Medicare and Medicaid Services. (2018). *National Healthcare Expenditures Fact Sheet*. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet.html

Daya, Z., & Hearn, J. H. (2018). Mindfulness interventions in medical education: A systematic review of their impact on medical student stress, depression, fatigue and burnout. *Medical Teacher, 40*(2), 146–153.

Dewa, C. S., Loong, D., Bonato, S., & Trojanowski, L. (2017). The relationship between physician burnout and quality of healthcare in terms of safety and acceptability: A systematic review. *British Medical Journal Open, 7*, e015141. https://doi.org/10.1136/bmjopen-2016-015141

Dobin, P. L., & Hutchinson, T. A. (2013). Teaching mindfulness in medical school: Where are we now and where are we going? *Medical Education, 47*(8), 768–779.

Dyrbye, L. N., West, C. P., Satele, D., Boone, S., Tan, L., Sloan, J., et al. (2014). Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. *Academic Medicine, 89*(3), 443–451.

Eley, D. S., Leun, J., Hong, B. A., Cloninger, K. M., & Cloninger, C. R. (2016). Identifying the dominant personality profiles in medical students: Implications for their Well-being and resilience. *PLoS One, 11*(8), e0160028.

Fridner, A., Fridner, A., Belkić, K., Marini, M., GustafssonSendén, M., & Schenck-Gustafsson, K. (2012). Why don't academic physicians seek needed professional help for psychological distress? *Swiss Medical Weekly, 142*, w13626. https://doi.org/10.4414/smw.2012.13626

Gabbe, S. G., Melville, J., Mandel, L., & Walker, E. (2002). Burnout in chairs of obstetrics and gynecology: Diagnosis, treatment, and prevention. *American Journal of Obstetrics and Gynecology, 186*(4), 601–612.

Goldstein, S. M., & Ward, P. T. (2004). Performance effects of physicians' involvement in hospital strategic decisions. *Journal of Service Research, 6*(4), 361–372.

Goodman, M. J., & Schorling, J. B. (2012). A mindfulness course decreases burnout and improves well-being among healthcare providers. *International Journal of Psychiatry in Medicine, 43*, 119.

Gorenstein, D. (2017). *Doctors are arming themselves with MBAs to navigate a tricky health care landscape*. Marketplace, Health. https://www.marketplace.org/2017/02/15/health-care/why-doctors-are-arming-themselves-mbas-navigate-tricky-health-care-landscape

Hojat, M. A. D., Spandorfer, J., & Mangione, S. (2013). Enhancing and sustaining empathy in medical students. *Medical Teacher, 35*, 996–1001.

Jain, S., & Miller, E. (2012). Why coursework in health policy and systems should be a premedical admission requirement. *Academic Medicine, 87*(5), 550–551.

Krasner, M. S., Epstein, R. M., Beckman, H., Suchman, A. L., Chapman, B., Mooney, C. J., et al. (2009). Association of an educational program in mindful communication with burnout, empathy, and attitudes among primary care physicians. *Journal of the American Medical Association, 302*(12), 1284–1293. https://doi.org/10.1001/jama.2009.1384

Lathrop, D. (2017). Disenfranchised grief and physician burnout. *Annals of Family Medicine, 15*(4), 375–378.

Lemaire, J. B., & Wallace, J. E. (2014). How physicians identify with predetermined personalities and link to perceived performance and wellness outcomes: A cross-sectional study. *BMC Health Services Research, 14*, 616.

Levine, L. S., & Gustave, L. (2013). Aligning incentives in health care: Physician practice and health system partnership. *Clinical Orthopaedics and Related Research, 47*(6), 1824–1831.

Lipsenthal, L. (2005). The physician personality: Confronting our perfectionism and social isolation. *Holistic Primary Care, 6*(3), 19.

Ljuboja, D., Powers, B. W., Robbins, B., Huckman, R., Yeshwant, K., & Jain, S. H. (2016). When doctors go to business school: Career choices of physician-MBAs. *The American Journal of Managed Care, 22*(6), e196–e198.

Mou, D., Sharma, A., Sethi, R., & Merryman, R. (2011). The state of health policy education in U.S. Medical Schools. *New England Journal of Medicine, 364*, e19. https://doi.org/10.1056/NEJMp1101603

Privitera, M. R., Rosenstein, A. H., Plessow, F., & LoCastro, T. M. (2015). Physician burnout and occupational stress: An inconvenient truth with unintended consequences. *Journal of Hospital Administration, 4*(1), 1–8.

Reynolds, N. T. (2012). Disruptive physician behavior: Use and misuse of the label. *Journal of Medical Regulation, 98*(1), 8–19.

Rosen, I. M., Gimotty, P. A., Shea, J. A., & Bellini, L. M. (2006). Evolution of sleep quantity, sleep deprivation, mood disturbances, empathy, and burnout among interns. *Academic Medicine, 81*(1), 82–85.

Rosenstein, A. H. (2015). Physician disruptive behaviors: Five year progress report. *World Journal of Clinical Cases, 3*(11), 930–934.

Rosenstein, A. H. (2017). Physician dissatisfaction, stress, and burnout, and their impact on patient care. In P. Papadakos & S. Bertman (Eds.), *Distracted Doctoring*. New York: Springer.

Schernhammer, E. S., & Colditz, G. A. (2004). Suicide rates among physicians: A quantitative and gender assessment (meta-analysis). *American Journal of Psychiatry, 161*(12), 2295–2302.

Shanafelt, T. D., Hasan, O., Dyrbye, L. N., Sinsky, C., Satele, D., Sloan, J., et al. (2015). Changes in burnout and satisfaction with work-life balance in physicians and the general US working population between 2011 and 2014. *Mayo Clinic Proceedings, 90*(12), 1600–1613.

Shanafelt, T. D., Sloan, J. A., & Habermann, T. M. (2003). The well-being of physicians. *American Journal of Medicine, 114*(6), 513–519.

Sinsky, C., Colligan, L., Li, L., Prgomet, M., Reynolds, S., Goeders, L., et al. (2016). Allocation of physician time in ambulatory practice: A time and motion study in 4 specialties. *Annals of Internal Medicine, 165*, 753–760.

Spickard Jr., A., Gabbe, S. G., & Christensen, J. F. (2002). Mid-career burnout in generalist and specialist physicians. *American Medical Association, 288*(12), 1447–1450.

Thomas-Dyrbye, L., & Shanafelt, T. D. (2006). Systematic review of depression, anxiety and other indicators of psychological distress among US and Canadian medical students. *Academic Medicine, 81*(4), 354–373.

West, C. P., Shanafelt, T. D., & Kolars, J. C. (2011). Quality of life, burnout, educational debt, and medical knowledge among internal medicine residents. *Journal of the American Medical Association, 306*, 952–960.

Woolhandler, S., Campbell, T., & Himmelstein, D. U. (2003). Costs of health care administration in the United States and Canada. *New England Journal of Medicine, 336*(11), 769–774.

# Chapter 22
# Bias, Disguise, and Co-opted Science: Altruism as "Scientized" Ideology Across the English Professions—The Peculiar Case of "Ebonics"

**Bradley Harris**

It is widely recognized that Southern American Englishes, taken collectively, are the most widely studied dialect group across the English language. The dialects of Black Americans have generally been considered either variants of Southern English or historically rooted in Southern English, in consequence of the northern migrations of Black Americans during several periods since the Civil War and their continuing contact with relatives in the South since. The speech of Southern and urban blacks, often loosely considered a "sub-dialect" (or sub-dialects) of Southern English, has gone by a variety of names: Black English, American Black English, and, more recently, African American Vernacular English (AAVE). At times, such dialects have borne labels which hover between being descriptive for linguists and judgmental for those outside the linguistic community. The descriptor "Non-Standard Negro English," (Rickford, 2019) seen in the 1960s, comes to mind as an example of a term which vibrates somewhere near a midpoint between those poles.

To the modern professional linguist, the term *non-standard* simply means a grammatical, lexical, or pronunciative form which, as a matter of fact, is outside the generally accepted standard or preferred form for the language. Nearly everyone's speech is non-standard in some way or other. To those outside the linguist's arena, however, the term *non-standard* can—quite understandably—carry a decidedly negative judgmental flavor. In the popular imagination, *non-standard* quickly becomes *sub-standard*. And so long as the "standard," in any realm, is the pinnacle, the ideal, that which is non-standard is indeed literally sub-standard: everything which is not at the North Pole is south of it.

This chapter examines ways in which science's principles, tools, and specific findings have been co-opted by various fields within the English professions, especially linguistics and the teaching of language. Appeals to scientific authority have often underlain efforts to claim altruistic or noble purpose. Especially when linguistics is translated for popular audiences or applied purposes, some such appeals have

B. Harris (✉)
Rhodes College, Memphis, TN, USA



D. M. Allen, J. W. Howell (eds.), *Groupthink in Science*,
https://doi.org/10.1007/978-3-030-36822-7_22

involved science oversimplified, misrepresented, truncated, or distorted. Nowhere is this more evident than in claims surrounding the concept of linguistic relativism, around which an entire culture of teaching has been built. The very concept of a *language* also derives from a long-standing "scientized" tradition in historical linguistics, but now is experiencing a deconstruction as that tradition is questioned. Imported into one or more fields within the English profession, such snatches of science or the literally and figuratively capitalized *Scientific Method* become supports for theories and programs which reach beyond the journals and into the practical world, flying the flag of "scientific" legitimacy but betraying underlying ideology. The term *ebonics*, and some of the values surrounding it, give us a fine focus for examination.

The ebonics controversy was recently described in a popular work from Oxford University Press thus:

> [D]ebates were sparked when the board of a school in Oakland, California, voted to change its policy regarding the education of African American children in Standard English. Given their consistently low level of achievement in the standard language, the board resolved to extend greater recognition to the vernacular spoken by the children themselves—a variety known to scholars as African American English (AAE), and more widely as Ebonics (Horobin, 2016, p. 78).

It is less than clear that the name *ebonics* had become "widely" known at the time. The term had been coined by social psychologist Professor Robert L. Williams of Washington University in St. Louis in the 1970s (Williams, 1975) as a contraction of *ebony* and *phonics*. Williams emphasized the international aspect of the dialect and its historical origins in the slave trade and the circumstances of slaves' living conditions, not only in the United States but in the Caribbean and other Western slave-holding nations. Williams's original definition focuses importantly—as the contraction *ebonics* itself implies, to quote Williams himself—on the "science of black speech sounds and language" (Quoted in Baugh, Baugh, 2019). Attention was thus drawn to such historical factors as the inferior educational opportunities afforded to slaves, and to African Americans generally, during and since the time of slavery. African Americans simply have not largely had the same opportunities to learn standard forms of American English as have people of European ancestry. The speech of many African Americans is, no doubt, *different* from other dialects, and different from what almost anyone would call "standard American English."

Are we to consider *ebonics*, then—if we are to consider the term an appropriate one and are to use it at all—a dialect or a language? Gloria Toliver-Weddington (1979) would have it that *ebonics* is a fair term and that ebonics is a dialect of English. More specifically, she claims ebonics is what had by the late 1970s long been called Black English (and now would be by most linguists termed African American Vernacular English or AAVE) (Toliver-Weddington, 1979). Two concerns with Toliver-Weddington's account are (1) that her concern is with education, "applied linguistics," if you will, and not linguistics in the scientific sense, and (2) that she does not consider at any great length the meaning(s) of the term *dialect*. Even before we get to these, however, there is the more basic question of *what*, precisely, Toliver-Weddington would say the term *ebonics* encompasses. Does it

embrace the broad and international set of speech forms to which Williams applies the term? Or does her use of *ebonics* encompass more narrowly American forms of speech? The answer is less than clear from her work. Developments since, in the theory of language and dialect, have not made the question any easier to deal with.

John Baugh rightly acknowledges the painful history of English as spoken by many African Americans. Both before and after the repeal of slavery, and ever since, he observes:

> A recurrent combination of racial segregation and inferior educational opportunities prevented many African Americans from adopting speech patterns associated with Americans of European ancestry…[G]enerations of white citizens maligned or mocked speakers of AAVE, casting doubt on their intelligence and making their distinctive speaking patterns the object of racist ridicule (Baugh, 2019).

To intrude a personal observation, I have noted, in teaching literature, writing, linguistics, and public speaking at colleges and universities in the Memphis area, such derision is sometimes so pervasive that even African American students will voluntarily apply labels such as *ignorant* or *stupid* even to specific speech patterns they themselves use, such as/aks/for *ask*,/errbody/for *everybody*, and *or either…* in place of standard *either…or*. Equally, I have, in a quarter century of living in Memphis and teaching English here, I have become quite accustomed to the ready willingness of many white Southerners to see in our shared skin color a presumed will to share in such ridicule. It has been observed by others that "the most stinging scorn for African-American mass culture is often expressed by middle-class African-Americans" (Hitchings, 2011, p. 257).

The linguistic effect of this racial clash was further problematized in 1996, in Oakland, California, when a resolution of the Oakland School Board created what Henry Hitchings called the "greatest American linguistic controversy of the last century" (Hitchings, 2011, p. 257). The resolution directed that African American students be instructed in "their primary language"—namely, for the Board, ebonics. What was important, however, was the perceived elevation of those students' brand of speech to the rank of language. Not just a narrow or local style of colloquiality. Not a mere dialect. No, this time, it was a language. Its origin, said the resolution, lay in "West African and Niger-Congo African language."

A professional linguist would have begun reply, perhaps, by insisting upon saying *languages*, plural, and pointing out that many tens, even hundreds, of languages of several families likely were involved in the linguistic origins of ebonics, including many beyond the West African and Niger-Congo regions and language groups. That linguist would likely also have pointed out that several competing theories vied for position in describing the roles of and relations between English and the slaves' languages of origin, as well as languages and dialects they encountered along the way from Africa to final destinations in America. These historical linguistic issues were not, however, the concern of the Oakland Board.

The Board sought legitimation. It wanted something for its students to stand on. Those students could not stand, it was clear, on a platform of "bad English"—a platform of indignity. The Board could not say, simply: *Our students speak*

*non-standard English. We are going to teach them standard English*. As the Board reasoned, they might as well have said "*Our students are broken, and wrong. We are going to fix them*." The Board, by granting to their African American students' speech the legitimacy of the brand *language*, sought to place them on an equal footing with those—largely white—students who had been able to claim the legitimacy of claiming to speak something nearer to a standard American English.

The nation was not happy. As Henry Hitchings observes, "[p]lenty of loud and poorly informed commentators" thrashed about in the popular press and semi-popular magazines, objecting strenuously to the Oakland Board's decision. At the core of their outcries was the objection that ebonics—whatever range of speech forms that term may have referred to, exactly—should not be legitimated as a dialect, let alone a language, "but simply as a corrupt and base type of English" (Hitchings, 2011, p. 257). Hitchings cites as especially vitriolic a *New York Times* piece faulting "theorists, lushly paid consultants, and textbook writers all poised to spread the gospel…that 'time that should be spent on reading and algebra [get] spent giving high fives and chattering away in street language'" (Hitchings, 2011, p. 258). To many in a population largely educated in the grammarian tradition of "proper English," the Board's legitimizing non-standard African American speech seemed a horror. Then as now, the error seemed to these objectors nothing less than a *moral* mistake. Hitchings' analysis helps explain why. He couches his account in terms of the growing English-only movement, which was already under full steam at the time of the Oakland incident.

The United States has no legislation specifying a single official language at the federal level. However, at least 31 of its states have legislation specifying English as an official language. Of these, Hawaii and Alaska have also specified one or more other languages as official. All the five inhabited U.S. territories have specified English as an official language.

Four of these have specified other official languages as well.[1] Except for Puerto Rico's Spanish, all the non-English specified official languages of U.S. states and territories are languages native to those areas. In total, then 36 U.S. jurisdictions have English as their official language and, of these, 30 have *only* English as an official language. The intent to make English official is clearly well underway. It is less than clear, however, that the English language is under threat of disappearance in America.

To speak speculatively, what may be the case is that the language is, in the minds of "standard American English" speakers, under threat of losing whatever degree of purity or correctness it may have. It is a commonplace observation, and has been for many decades, that English is "in decline." Thus, *any* departure from the bygone rules of revered and reimagined high school English teachers is to be lamented.

One does not have to be racist per se to oppose the legitimation of dialects such as those of black youth, whether termed ebonics or African American Vernacular

[1] "English Only Movement." Wikipedia. https://en.wikipedia.org/wiki/English-only_movement. Retrieved 01 July 2019.

English. One simply must be rigid. To oppose all departure from grammatical tradition—"splitting the infinitive was wrong when I was in school, and it's wrong now"—involves no necessary racial or xenophobic component. Such insistence involves only resistance to linguistic change from outside *linguistic* forces.

That said, it is clear that the "English only" movement, as framed and advanced by such figures as Theodore Roosevelt and S.I. Hayakawa, is very much directed at the consolidation of single national language as part of a unified national culture" (Hitchings, 2011). Roosevelt's century-old one flag, one language vision of America is very much alive. It is the meat and potatoes of English-only activists today. It is common fare, also, among the rank and file of citizens. Listen to any morning's or afternoon's worth of talk radio, and you are bound to hear some recitation of the notion, "*If they want to come here, they'd better learn English fast*…".

What is more, it had better be some brand of English the rest of us—white people—can readily understand. This latter demand we may take as a call for something like Standard English, or Standard American English, or as some prefer, American.

But what is that? Quirk and Greenbaum's definitive *Comprehensive Grammar* notes that, in affirming

> Students' right to their own varieties of language, many American educationalists have declared that Standard American English is a myth, some asserting the independent status of (for example) Black English. At the same time, they have acknowledged the existence of a written standard dialect, sometimes termed 'Edited American English' (Quirk, Greenbaum, Leech, & Svartvik, 2010, p. 20).

Wisely, Quirk, Greenbaum et al. have acknowledged that it is less than clear what Standard American English is. The 1800-odd pages of their grammar recognize again and again in one important sense the supremacy of dialect over language. Over and over, they note that one feature or form is acknowledged correct by speakers of one dialect but not by those of another. (One simple example: the tendency of Americans to call the alphabet's last letter *zee*, while Canadians often call it *zed*.)

We can certainly see African American Vernacular English as a dialect. Linguist William Labov prefers to see it as "a subsystem of English" with its own phonological and syntactic rules…now aligned…with rules of other dialects." He sees AAVE as both incorporating features of Southern English and as having affected Southern English. Labov is a creolist, seeing AAVE as having grown from an earlier creole similar to those of the Caribbean. Finally, AAVE has a highly developed verb-aspect system showing continuing growth of its semantic structure (Mufwene, 2001). As Seth Lerer Lerer (2007) points out, to find the full measure of distinction and substance in AAVE, we would have to look beyond the mechanics of phonology, morphology, lexicon, syntax, semantics, and the like. We must further acknowledge that AAVE is not spoken by all African Americans, that not all of the dialect's speakers are African American, and that AAVE may not be a unified dialect. History, rhetoric, theatre, and other disciplines must inform what we are to discover, as observed by Henry Louis Gates (1988).

Altruism, in its strict or literal sense, may be hard enough to find anywhere. However, it is not difficult, given a modicum of sympathy, to find a solid measure good will, good intentions at least, within the Oakland Board of Education and the educators who gave rise to their 1996 resolution, however misguided, however ill-founded it may have been scientifically and linguistically. Similarly, let us suggest also that opposing positions, however unexamined some of them may be, are not without their elements of goodwill.

What we see throughout the political side of the discussion, however—the portion occurring outside the community of professional, scientifically oriented linguists—is repeated misconstrual of concepts. Notions such as *language* and *dialect* are misconceived and misapplied. Existing definitions are ignored along with already acknowledged difficulties in definition. Also ignored were the likely natures of opposing arguments.

The linguistic community itself has experienced a long period of unsettlement on issues relevant to this discussion, and only comparatively recently has come to any degree of consensus amid discussions of the complex nature of African American Vernacular English. AAVE may not be a single dialect—hence Labov's term *subsystem of English*—and *dialect* may not be the best term for this set of speech forms. Whatever else be true, it does seem that *ebonics*—with or without the capital E Williams originally employed with the term—is not a felicitous term. Nor did it, nor will it likely ever lead to desirable results.

## References

Baugh, J. *Comprehending Ebonics*. www.pbs.org/speak/seatosea/americanvarieties/AAVE/ebonics/. Retrieved 01 July 2019.

Gates, H. L. (1988). *The signifying monkey* (pp. 72–75). New York: Oxford University Press.

Hitchings, H. (2011). *The language wars: A history of proper English*. New York: Farrar, Strauss & Giroux.

Horobin, S. (2016). *How English became English* (p. 78). Oxford, UK: OUP.

Lerer, S. (2007). *Inventing English: A portable history of the language* (pp. 220–234). New York: Columbia University Press.

Mufwene, S. S. (2001). African-American English. In J. Algeo (Ed.), *The Cambridge History of the English Language* (*English in North America*) (Vol. 6, pp. 291–324). Cambridge, MA: CUP.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (2010). *A comprehensive grammar of the English language*. New Delhi, India: Pearson, §1.24, p.20.

Rickford, J. R. (2019). What is Ebonics (African American English)? *Linguistic Society of America*. www.linguisticsociety.org/content/what-ebonics-african-american-english. Retrieved 02 January 2019.

Toliver-Weddington, G. (1979). Ebonics (Black English): Implications for education. *Journal of Black Studies, 9* (No. 4) [special issue].

Williams, R. (1975). *Ebonics, the true language of black folks*. St. Louis, MO: Robert Williams and Associates.

# Index




D. M. Allen, J. W. Howell (eds.), *Groupthink in Science*,
https://doi.org/10.1007/978-3-030-36822-7