The Great Debate, Revisited

The Great Debate was a pivotal event in the history of biobehavioral medicine, or at least it should have been. It took place in Monterey, California, on March 9, 2001, at the 59th Annual Meeting of the American Psychosomatic Society. It was organized by Jerome H. Markovitz, MD, MPH, Associate Professor of Medicine at the University of Alabama at Birmingham, with the assistance of one of the authors of this paper (G.E.M.). Tragically, Dr. Markovitz died of pancreatic cancer on September 5, 2002, at the age of 45. Those of us who had the privilege of knowing Jerry as a friend and colleague miss him dearly. This article is dedicated to him.

As we approach the 5th anniversary of this extraordinary event, we are pausing to reflect on what could and should have been learned from it. We feel compelled to do so because little has been written about it in the 5 years that have elapsed since it took place. We could have devoted a Statistical Corner to it instead of an editorial, because much of the debate focused on our research methods and on how we interpret (or misinterpret) our findings. We also could have consigned it to the Existential Corner, if only we had one, because the Great Debate raised fundamental questions about our raison d’être and our future as a field of scientific research.

Interested readers may wish to review the redacted transcripts of the debate that were published in this journal (1–5). An audio recording of the debate is also available at www.psychosomatic.org/ed_res/index.htm.1 It is well worth hearing. It reveals some of the heat and humor that written transcripts cannot convey, and it adds nuance to the arguments that swept ashore on that memorable day along the beautiful coast of northern California.

An Overview of the Debate

“The resolution was presented by two of the leading figures in biobehavioral medicine: Redford B. Williams, MD, from the Department of Psychiatry at Duke University Medical Center in Durham, North Carolina, and Neil Schneiderman, PhD, from the Behavioral Medicine Research Center at the University of Miami in Coral Gables, Florida. Dr. Williams presented research on cardiovascular disease, and Dr. Schneiderman addressed cancer and HIV/AIDS. They argued that 1) psychosocial factors influence the development and course of these medical illnesses via biologically plausible mechanisms, 2) there are efficacious interventions for these psychosocial problems, and 3) these interventions also improve medical outcomes. They advanced these arguments by presenting evidence from 23 articles, including 11 interventional and 12 observational studies, that they regarded as being among the best examples of biobehavioral research available at the time. Their opponents added one more interventional article to the list, for a total of 24. Five of the papers had been published in Psychosomatic Medicine; the rest had appeared in New England Journal of Medicine, JAMA, Lancet, or other respected journals.

The case against the resolution was presented by two former Editors-in-Chief of New England Journal of Medicine: Arnold Relman, MD, was a Professor Emeritus of Medicine and Social Medicine at Harvard University Medical School, and Marcia Angell, MD, was a Senior Lecturer in Social Medicine at Harvard University Medical School. The fact that two of the leading figures in biobehavioral medicine faced such formidable opponents helps to explain why this event was called “The Great Debate.”

After the affirmative speakers presented their case, Dr. Relman began, remarkably, by conceding that, “(t)he mechanisms by which brain and mind interact with the body may be debatable, but the fact of the connection is established and we do not doubt that.” He followed, however, by stipulating that the debate should concern only direct, physiologic mechanisms, not behavioral ones: “Given the fact of the mind-body connection, is there any good clinical evidence that psychological and social interventions can directly change the course of serious organic disease? By directly, we mean through some direct effect on the biological process itself rather than through changes in compliance with treatment or some change in behavior such as diet or exercise that might well affect the course of illness.”

These remarks set the stage for a number of subsequent exchanges about behavioral mechanisms. Dr. Schneiderman spent a substantial portion of his rebuttal arguing that behavioral or lifestyle factors play crucial roles in health and disease, but his opponents did not dispute this point. In fact, they readily acknowledged it. They made it quite clear, however, that they were not there to discuss health behavior. They had come only to debate whether various forms of emotion dysregulation such as anger, depression, or psychological stress have direct, physiologic effects on medical illnesses such as coronary heart disease, whether psychosocial or behavioral interventions for these emotional problems can improve med-

1RealPlayer software is required in order to listen to the recording. A free version of the software may be downloaded from www.real.com.

Address correspondence and reprint requests to Kenneth E. Freedland, PhD, Department of Psychiatry, 4625 Lindell Blvd., Suite 420, St. Louis, MO 63108. E-mail: freedlak@bmc.wustl.edu

DOI: 10.1097/01.psy.0000209374.25201.4c
The negative speakers divided the articles in question not by disease but instead by design; Dr. Relman discussed the intervention studies, and Dr. Angell discussed the observational studies. Both were harshly critical of the papers they reviewed. Dr. Relman, for example, asked, “(h)ow good is the evidence in those 23 articles? After examining them carefully, it is our contention that none of them provides good strong evidence. Most of them, in fact, are so flawed as to be uninterpretable.” He went on to say that the interventional studies suffered from serious weaknesses such as inadequate data on possible behavioral differences between treatment and control groups, inadequate adjustment for other potential confounders, small samples, lack of blinding, use of subjective endpoints, and failure to follow the intention-to-treat principle in outcome data analyses. Dr. Relman was not, it should be recalled, referring to the worst articles that could be dredged up, or even to a random assortment of typical studies. He was talking about a set of papers, hand-picked by two of the leading researchers in our field, which comprised some of the best examples of biobehavioral science circa 2001.

Dr. Angell’s critique was scathing. She noted that, “(o)bservational epidemiologic studies are fraught with difficulties, and these articles exemplify most of them. The most serious are: 1) failure to deal adequately with possible confounding variables, 2) failure to distinguish cause from effect, 3) data dredging, and 4) biased interpretation.” Later, she stated that, “I have spent the last 21 years reviewing and editing many scientific papers, and it strikes me that the literature on psychosomatic interventions and associations is unusually poor. In general, papers on this subject are not as rigorous as those in other areas. There seems to be a double standard.”

Dr. Angell was the first panelist to claim the existence of a double standard, but she was not the last. In their closing arguments, the affirmative side spent more time on this issue than any other. They argued that their opponents were the ones who held a double standard. In short, while Drs. Relman and Angell asserted that the methodological standards of research in biobehavioral medicine are too low, Drs. Schneiderman and Williams countered that their opponents’ demands for methodological rigor are too high.

Drs. Williams and Schneiderman conceded that all of the studies they had presented had flaws and limitations, but they insisted that none of them were fatally flawed and that the weight of evidence from these studies supported the resolution. In their closing arguments, Drs. Relman and Angell vehemently rejected the charge that they hold psychosocial intervention trials or other biobehavioral studies to higher methodological standards than the ones they apply to mainstream medical research. They also argued that the evidence in support of the resolution did not weigh very much after all. Dr. Relman noted, “(y)ou cannot strengthen your argument, Dr. Schneiderman, simply by accumulating a lot of studies that are not very good. A large number of weak studies do not add up to a strong conclusion. To argue that these papers we criticized, although not as good as one might like, did not have truly fatal flaws, and therefore in total ought to be considered as valid evidence—that is not good science or even rational thinking. We must be driven by the evidence and the evidence must be credible.”

The debate was a plenary session; almost everyone at the conference attended it. The audience included some of the most eminent researchers in our field, some of our most rapidly rising stars, and some of the principal investigators of those notoriously weak studies that apparently did not add up to very much. After hearing the closing arguments, the audience might have wanted to go down to the beach and drown itself en masse, but everyone stayed to find out what the discussant had to say. He was, after all, no less an authority on medical research than George Lundberg, MD. Dr. Lundberg had been the Editor-in-Chief of the Journal of the American Medical Association for 17 years before becoming the Editor-in-Chief of MedScape. He had also been a distinguished Professor of Pathology at the University of Southern California and Chair of Pathology at the University of California at Davis.

Anyone hoping to be comforted after listening to the debate must have been reassured by Dr. Lundberg’s gracious introduction and his claim that he was completely unbiased because “I know virtually nothing about the subject.” It soon became apparent, however, that he did know something about the subject. In fact, he first learned about the physiological effects of stress in 1955, as a medical student, directly from Hans Seyle, the renowned pioneer of stress research. He recalled Seyle having said that “…stress was everywhere, it affected all kinds of bodily functions in major ways and that different people responded very differently.” He added, “That was in 1955, and I think that is about where we are (no offense intended) in 2001.” He went on to acknowledge that social and psychological interventions may be useful in the primary and secondary prevention of medical illnesses, and that they may help to improve the quality of life of patients with chronic diseases. He did not, however, offer any support for the proposition that such interventions can affect hard medical outcomes in cardiovascular disease, cancer, AIDS, or any other form of physical illness.

Who Won the Debate?

President Kennedy once said, “Victory has a thousand fathers, but defeat is an orphan.” Regardless of which side won the debate on points, we in the biobehavioral research community abandoned it like an orphan on the doorstep of medical science. Had the affirmative side decisively won, we would not have so quickly forgotten it. They did not decisively win, but their double-standard defense did help us forget. It struck a deep chord among those of us who believed that our research was not being taken seriously enough by the medical mainstream. It consoled us, too: We had no reason to worry about what such implacable critics thought of our best work; they were just holding us to a double standard. It was as though we had collectively responded to a cascade of criticism.
THE GREAT DEBATE, REVISITED

by telling ourselves, like Al Franken’s comic character Stuart Smalley, “We’re good enough, we’re smart enough, and dog-gone it, people like us.”

What if we were to swallow our pride and take Drs. Relman and Angell at their word? What if we were to give them the benefit of the doubt and accept their assertion that they hold our research only to the same high standards to which they hold everyone else’s? What if we were to acknowledge that despite all of the progress we have made, they are not alone among tough-minded medical scientists in having reasonable doubts about the strength of our evidence (e.g., 6)? What could we learn from this dolorous thought experiment? In a word, plenty.

To See Our Work as Critics See It

Our critics doubt that any form of emotional dysregulation is a proven causal risk factor for any form of medical morbidity or mortality. It is encouraging that they do not require extraordinary proof, since to them, ours is apparently an extraordinary claim: Like the mental equivalent of an autoimmune disorder, our own emotions can make us sick.

For modern medical science to flourish, it had to overcome ancient superstitions about disease, such as the notions that sickness results from ill humors or demon possession. To their chagrin, contemporary biomedical scientists who are dedicated to building on their predecessors’ immense legacy of progress are facing a resurgence of superstitious and pseudoscientific beliefs about disease, and of snake oil remedies for sale to those who hold them. We vehemently reject any attempt to lump serious biobehavioral research on the health effects of emotional dysregulation together with this anachronistic drivel. We would do well to recall, however, that Drs. Relman and Angell did not do so. To the contrary, they took our proposition quite seriously. They disputed the evidence for it, but they also assured us that they thought convincing evidence could eventually accrue, as long as the phenomena in question truly exist and our research methods are truly rigorous.

What, exactly, are the phenomena in question? The pathogenicity of multiple types of emotional dysregulation, including stress, anger, depression, anxiety, and others, are under active investigation. There are multiple variants of each, and multiple intersections among them. Many of us tend to emphasize certain forms or combinations in our own research, and to neglect or reject others. Some of us have also attempted to demonstrate the dominance of our own favorite forms over others. Clearly, though, our critics do not care which ones we dare to proclaim. It’s all the same to them, and they are indifferent to our internecine conflicts.

We cannot be indifferent, however, since we do not yet know which forms, variants, or combinations of dysregulated emotions (if any) cause medical morbidity or mortality. We may be defining our emotional phenotypes too broadly or too narrowly, and we may be measuring them too imprecisely. The possibilities grow exponentially when we consider that the bad actors may differ among conditions as diverse as cancer, AIDS, and coronary disease, and across different stages of these conditions. We are convinced that some fairly sharp needles are buried in these haystacks, but the predictors we have examined so far have been like handfuls of hay. Some of them hold needles, we think, but some of them may not. Dr. Angell contended that we have been studying “weak effects” that are “easily swamped by effects of confounding variables.” Stronger and more consistently replicable effects would be much more difficult to dispute, so it would behoove us to find ourselves some needles and see where we can stick them.

The needles we need to prove our point must be threaded with a certain kind of sinew. Our critics will not be convinced, even by strong effects, unless we can prove that they are mediated by direct, physiologic mechanisms. They do not have to be the only pathways linking emotional dysregulation to medical morbidity or mortality, but they have to exist.

Or do they? Our critics think so, but are the effects of emotional dysregulation on disease truly any less important if they are mediated by behavioral mechanisms rather than by physiological ones? From an evolutionary perspective, for example, the biological alterations that accompany stress occur primarily to support behavioral responses, and they are in turn regulated by the organism’s behavioral responses to stressful challenges. But ever since the days of Cannon and Selye, psychosomatic scientists been much more enthralled by the physiology and fearfulness of stress than by the fighting, fleeing, or freezing it provokes (7). Our faith in physiology has grown apace since then; it animates our research on emotions and disease. Our critics simply want the proof that warrants our enchantment. Their demand did not spring forth from a vacuum; it grew out of the field that we ourselves have tilled. If physiology eludes us in the end, the disappointment will be ours, not theirs. But our field is hardly fallow; little physiologic shoots are sprouting up all over (e.g., 8–11). We just have to help them grow.

Clinical significance is also what’s at stake in this. Emotional dysregulation clearly affects health behavior; we know that depression, for example, doubles or triples the risk of nonadherence to medical treatment regimens (12,13). But it is not necessarily the predominant cause of nonadherence or of other dysfunctional health behaviors. Environmental contingencies, genetic predispositions, social factors, and beliefs collectively explain as much if not more of the variance in health behavior. In this light, if the health effects of dysregulated emotions were mediated exclusively by health behavior, the critics would have good reason to question our preoccupation with these emotions, and we would too. We would have to ask ourselves why emotional dysregulation deserves so much more of our scientific attention than we devote to the maladaptive behaviors they merely help promote. We should pay more attention to health behavior anyway, but the clinical importance of dysregulated emotions qua causes of disease inevitably hinges, to some degree, on whether any physiologic mechanisms are involved.

Unfortunately, the mechanistic research in this field has been fragmented and compartmentalized. There have been studies, for example, of physiologic mediators such as inflam-
matory cytokines, and studies of behavioral ones such as nonadherence, but little research on whether or how the behavioral mechanisms interact with the physiologic ones. This works against us because evidence for physiologic pathways is incomplete if their interrelationships with behavioral pathways remain poorly understood. It is also incomplete if we do not establish that they are independent of plausible confounders, including shared genetic factors (14,15). If we are ever going to convince the critics, our mechanistic research must eventually yield comprehensive causal models in which dysregulated emotions are linked to hard medical outcomes via physiological mechanisms, not just in isolation but in relation to any coexisting behavioral pathways, any shared genetic factors, and any other significant confounders.

Mechanistic issues confront us, not only in observational studies but in clinical trials as well. Large clinical trials tend to be rather Spartan affairs, and it is difficult to gain inclusion of mechanistic measures in trial protocols. It will be necessary to do so anyway, if we are ever to prove that the medical outcomes of our interventions are mediated, at least in part, by physiologic mechanisms and not only by behavior. This will be largely moot, however, until we have developed some highly efficacious interventions. We had little proof of that 5 years ago, and not much more proof now. The results of the ENRICHD trial, for example, were published about 2 years after the Great Debate (16). The intervention had no effect on the primary medical endpoint, reinfarction-free survival after an acute myocardial infarction, and quite modest effects on the targeted psychosocial risk factors, depression, and low perceived social support.

The main reason to conduct clinical trials is to develop interventions that yield meaningful clinical outcomes, but another is to experimentally test causal hypotheses about risk factors. Clinical trials are indeed experiments, but not in the sense of randomly assigning one group to have a risk factor and another group to be free of it. Interventions that have only weak effects on risk factors will subserve only weak tests of our causal hypotheses; conversely, powerful interventions will permit stronger tests. We need more efficacious interventions to advance both the welfare of our patients and the frontiers of our biobehavioral science.

We should also remember, though, that random assignment to chronic risk factor exposure is possible in animal studies. Cynomolgus monkeys, for example, have been subjected to chronic social stressors in studies of coronary atherosclerosis (e.g., 17–22) and Sprague-Dawley rats have been exposed to chronic mild stress in a rodent model of the cardiovascular effects of depression (e.g., 23–26). This research has shown unequivocally that stress promotes the development of atherosclerosis and has identified some of the mechanisms that may be responsible. No matter how much progress we achieve in developing efficacious treatments for human emotional dysregulation, animal studies will continue to play an essential role in biobehavioral hypothesis testing. Animal research was excluded from the Great Debate, but we should embrace it.

Double Standards Redux

The audience posed some excellent questions for the debaters after Dr. Lundberg’s closing remarks. One individual asked whether there should be a double standard, not with respect to the rigor of our research but to the kinds of methods we should be expected to use. He argued that psychosocial research on risk factors for medical illness is much more complex than, for example, medical research on purebred mice, that human subjects do not fit neatly into the purified samples that Dr. Angell prefers, and that psychosocial measures are necessarily noisier than biological ones. But Dr. Angell held her ground. Although it may be quite difficult to conduct rigorous psychosocial research, our critics clearly see that as our problem, not theirs.

At least we are not alone; all clinical investigators have to work within ethical and methodological constraints. The fact remains, though, that some methodological ideals are simply beyond our reach. Double blinding, for example, is usually impossible in randomized clinical trials of psychosocial interventions. What should we do instead? An expert consensus panel recently developed a set of guidelines for assessing the quality of randomized controlled trials of nonpharmacological treatments (27). If we follow them, our trials will be as rigorous as they can possibly be, but will they be rigorous enough to satisfy our critics? Dr. Relman offered some encouragement on this point, but only time will tell whether studies that meet the emerging standards of nonpharmacological trials will ever convince clinical scientists who are accustomed to the standards of pharmacological trials. Despite the extraordinary methodological challenges that pervade human biobehavioral research, we have to abide by the same fundamental rules of evidence as they do. Ultimately, nothing less will be persuasive.

A related double standard inhered in the premise of the debate: “Resolved: Psychosocial interventions can improve clinical outcomes in organic disease.” As Dr. Schneiderman pointed out, most of the trials in question had tested behavioral rather than psychosocial interventions. Regardless, none of the participants questioned whether the debate should have been limited to nonpharmacological trials. The central issues are whether any form of emotional dysregulation has a causal role in any serious medical outcome, and if so, whether there is anything we can do about it. Pharmacological interventions may complicate our pursuit of mechanistic models, but they will inevitably play an important role in the clinical management of emotional dysregulation. If antidepressants, anxiolytics, or other psychopharmacological agents can improve clinical outcomes in organic disease, and if they can do so via effects on depression, anxiety, or other forms of emotional dysregulation, then they can help us test our causal hypotheses. Pleiotropy complicates the use of drugs for causal hypothesis testing, but it does not preclude it altogether. Furthermore, psychosocial interventions are also pleiotropic. Thus, there is no reason to stipulate that only psychosocial interventions trials can yield
THE GREAT DEBATE, REVISITED

bona fide evidence that treating psychosocial risk factors improves medical outcomes.

Another member of the audience identified disparities in research financing as yet another kind of double standard. There was general agreement among the panelists that, despite the complexity of our work, much less money is available to pay for it than is devoted to pharmacological research and to other domains of biomedical science. However, Dr. Relman rejected the contention that methodological rigor can only be achieved by spending vast sums of money. He noted that, for example, research on intermediate medical outcomes can be both informative and much less expensive than studies in which the primary endpoints are mortality or serious medical morbidities, as long as the conclusions are limited accordingly. Whether there will ever be sufficient funds to pay for definitive trials was a question left unanswered.

A Defining Moment

We may not have realized it at the time, but we were swept away by a scientific tsunami that day in Monterey. We survived, but we are still out at sea. We need to do more than find our way back ashore; we need to climb to higher ground once we get there. Not merely up the dunes, not wandering the foothills: Let’s head for Mt. Olympus, and plant our flag on top. It’s a long way from Monterey, but we have a long way to go if we are ever going to win the Great Debate.

Excellent work has been done and impressive progress has been achieved over the past 5 years. Many exciting new findings have been published during this period. It would probably have helped our cause, for example, if the results of INTERHEART (28–30) had been available 5 years ago. The problem is that we are still at an early stage in the evolution of our science; and at the rate we are going, the great questions raised in the Great Debate will not be definitively answered in the next 5 years, or even in the next 50. We should strive to answer them sooner than 2056; they are much too important for us to be that patient.

In order to accelerate our efforts, we will have to confront our own doubts. It may be that emotional dysregulation does not play a direct, physiologic, causal role in any major medical outcome; neither we nor our critics know for sure. Although some of them may believe that we are chasing the psychosocial equivalent of cold fusion and will fail in the end, we are inspired by the growing evidence for the emotional dysregulation hypothesis, even if it is still flawed and incomplete, and believe that we will eventually succeed. We need to remind ourselves that this is a very important goal, one that is well worth pursuing, and that it is up to us to reach it.

The burning question of our time is not whether but how to accelerate our efforts. Our collective scientific output is already impressive in terms of field-initiated research. Significant productivity growth in this arena can probably still be achieved if we remain committed to welcoming bright, well-trained, enthusiastic new investigators to join our quest. Nevertheless, our individual efforts have not been as well coordinated as they will have to be for us to reach our goal in the foreseeable future. Larger enterprises, such as the ENRICHD trial and NIH’s Mind-Body Centers initiative, have been exceedingly helpful, but even they pale in comparison to the immensity of the challenge that we have accepted.

The challenge is indeed immense. The comprehensive causal models that we are ultimately trying to establish comprise numerous components, each of which will be challenging in itself to define. This will require 1) more precise specification of the “toxic” forms of emotional dysregulation and of the critical periods during which they exert their effects; 2) conclusive evidence of their independence from important confounders; 3) determination of which medical outcomes, in which illnesses are affected by these forms of emotional dysregulation; 4) complete delineation of the biobehavioral pathways linking the emotions to these medical outcomes; 5) development of interventions that are highly efficacious in modifying the dysregulated emotions; 6) demonstration that these interventions also improve the medical outcomes in question; and 7) confirmation that their efficacy is at least partially mediated by the physiologic effects of these emotional gains.

A variety of methodologies, technologies, and disciplines will have to be brought to bear on each component, as well as on the system as a whole. Our task is less like building a structural equation model than entraining a recursive neural network; there will have to be both forward and backward propagation among the components and across studies. We cannot tell, for example, exactly which kinds of emotional dysregulation are the true bad actors until we know which medical outcomes they affect, and we cannot be sure which medical outcomes are affected until we know which kinds of emotional dysregulation affect them. Similarly, it is difficult to develop highly efficacious treatments for inadequately defined conditions, yet it is necessary to test such treatments in order to determine whether they affect medical outcomes.

All of this may seem like a gigantic Catch-22 and a recipe for paralysis, but it is both unnecessary and counterproductive to see it in such terms. Cardiovascular clinical trialists have learned, for example, to remain steadfast in their relentless pursuit of better medical outcomes, even while myriad mechanistic questions remain to be answered (31). We can do the same.

It will be a monumental challenge to obtain definitive evidence, the kind that will convince not only us but our critics that psychosocial interventions can improve outcomes in organic disease. We strongly believe, however, that this is achievable. Animal models have already shown that chronic stress and depression can have serious cardiovascular effects. Emerging technological and methodological advances are opening the physiologic floodgates in mechanistic research on the effects of emotional dysregulation on cardiovascular disease and other medical illnesses in humans.

In order for us to capitalize on these exciting new developments, many different experts will have to meet, collaborate, and focus sustained attention on a coherent set of long-term objectives. The requisite interdisciplinary collaborations extend beyond our own ranks to experts from related fields, and
substantial funding will be needed to make them possible. The funding will have to come from multiple sources and mechanisms. Our Roadmap to Olympus may well come from NIH, but we’ll also need much more than that, starting with a GPS device and a good, strong pair of boots.

The Great Debaters and their peers will have long since retired by the time we get there, but we owe it to the next generation of biobehavioral researchers to help them reach the top. It is just the 5th anniversary of the Great Debate. Let us resolve that by the 25th, they will win if there’s another.

KENNETH E. FREEDLAND, PhD,
Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri

GREGORY E. MILLER, PhD,
Department of Psychology, University of British Columbia, Vancouver, Canada

DAVID S. SHEPS, MD, MSPH
Department of Medicine, Division of Cardiovascular Medicine, University of Florida, Malcom Randall VA Medical Center, Gainesville, Florida

REFERENCES
