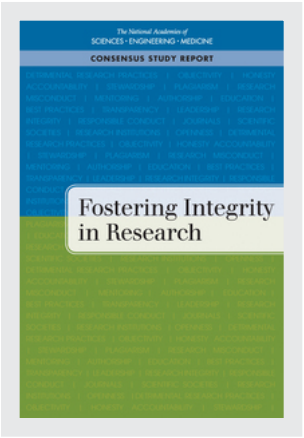


# Fostering Integrity in Research



## DETAILS

326 pages | 6 x 9 | PAPERBACK
ISBN 978-0-309-39125-2 | DOI 10.17226/21896

## CONTRIBUTORS

Committee on Responsible Science; Committee on Science, Engineering, Medicine, and Public Policy; Policy and Global Affairs; National Academies of Sciences, Engineering, and Medicine

# Fostering Integrity in Research

Committee on Responsible Science

Committee on Science, Engineering, Medicine, and Public Policy

Policy and Global Affairs

A Consensus Study Report of

*The National Academies of*
SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS
*Washington, DC*
**www.nap.edu**

**THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001**

This activity was supported by the Burroughs Wellcome Fund under Grant No. 1012589, the Office of Inspector General of the National Science Foundation under Contract No. NSFDACS11P1173, the Office of Research Integrity of the U.S. Department of Health and Human Services under Contract No. HHSP23320042509XI, the Office of Science of the U.S. Department of Energy under Contract No. DE-SC0005916, the U.S. Department of Veterans Affairs under Contract No. VA101-C17404, the U.S. Environmental Protection Agency under Contract Nos. EP-C-09-003 and EP-C-09-005, and the U.S. Geological Survey of the U.S. Department of the Interior under Contract No. G10AP00150, with additional support from the American Chemical Society, the American Physical Society, the Society for Neuroscience, the National Academy of Sciences Arthur L. Day Fund, and the W. K. Kellogg Foundation Fund. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project.

International Standard Book Number-13: 978-0-309-39125-2
International Standard Book Number-10: 0-309-39125-3
Library of Congress Control Number 2017947965
Digital Object Identifier: https://doi.org/10.17226/21896

Additional copies of this publication are available for sale from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.



Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2017. *Fostering Integrity in Research*. Washington, DC: The National Academies Press. doi: https://doi.org/10.17226/21896.

*The National Academies of*
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. C. D. Mote, Jr., is president.

The **National Academy of Medicine** (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.nationalacademies.org**.

*The National Academies of*
SCIENCES • ENGINEERING • MEDICINE

**Consensus Study Reports** published by the National Academies of Sciences, Engineering, and Medicine document the evidence-based consensus on the study's statement of task by an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and the committee's deliberations. Each report has been subjected to a rigorous and independent peer-review process and it represents the position of the National Academies on the statement of task.

**Proceedings** published by the National Academies of Sciences, Engineering, and Medicine chronicle the presentations and discussions at a workshop, symposium, or other event convened by the National Academies. The statements and opinions contained in proceedings are those of the participants and are not endorsed by other participants, the planning committee, or the National Academies.

For information about other products and activities of the National Academies, please visit www.nationalacademies.org/about/whatwedo.

## COMMITTEE ON RESPONSIBLE SCIENCE

ROBERT M. NEREM (*Chair*), Institute Professor and Parker H. Petit Professor Emeritus, Institute for Bioengineering and Bioscience, Georgia Institute of Technology
ANN M. ARVIN, Lucile Salter Packard Professor of Pediatrics and Professor of Microbiology and Immunology, Stanford University School of Medicine and Vice Provost and Dean of Research, Stanford University
REBECCA M. BERGMAN, President, Gustavus Adolphus College
MOSES H. CHAN, Evan Pugh Professor of Physics, Pennsylvania State University
C. K. GUNSALUS, Director, National Center for Professional and Research Ethics, Research Professor, Coordinated Science Laboratory, Professor Emerita, College of Business, University of Illinois at Urbana-Champaign
DEBORAH G. JOHNSON, Anne Shirley Carter Olsson Professor of Applied Ethics Emerita, University of Virginia
MICHAEL A. KELLER, Ida M. Green University Librarian, Publisher of Stanford University Press, Founder of HighWire Press, Stanford University
W. CARL LINEBERGER, E. U. Condon Distinguished Professor of Chemistry and Fellow of JILA, University of Colorado
BRIAN C. MARTINSON, Senior Research Investigator, HealthPartners Institute, Core Investigator, Minneapolis VA, Center for Chronic Disease Outcomes Research, and Associate Professor, University of Minnesota, Department of Medicine
VICTORIA STODDEN, Associate Professor of Information Sciences, University of Illinois at Urbana-Champaign
SARA E. WILSON, Associate Professor of Mechanical Engineering, University of Kansas
PAUL ROOT WOLPE, Asa Griggs Candler Professor of Bioethics, Director, Center for Ethics, Emory University
LEVI WOOD, Assistant Professor of Mechanical Engineering, Georgia Institute of Technology

### Staff

TOM ARRISON, Study Director
LIDA ANESTIDOU, Senior Program Officer
NINA WARD, Research Associate
NEERAJ GORKHALY, Research Associate (until February 2014)
MARIA LUND DAHLBERG, Associate Program Officer (until September 2016)
STEVE OLSON, Consultant Writer

**COMMITTEE ON SCIENCE, ENGINEERING, MEDICINE, AND PUBLIC POLICY**

RICHARD N. ZARE [NAS] (*Chair*), Marguerite Blake Wilbur Professor, Stanford University

CYNTHIA BARNHART [NAE], Chancellor, Ford Professor of Engineering, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology

DAVID E. DANIEL [NAE], Deputy Chancellor, The University of Texas System

VICTOR J. DZAU [NAM] (*ex-officio*), President, National Academy of Medicine

DIANE E. GRIFFIN [NAS, NAM], Distinguished University Service Professor, W. Harry Feinstone Department of Molecular Microbiology and Immunology, Johns Hopkins Bloomberg School of Public Health

JOHN G. HILDEBRAND [NAS], Regents Professor, Department of Neuroscience, University of Arizona

JUDITH KIMBLE [NAS], Investigator, Howard Hughes Medical Institute; Professor of Biochemistry and Molecular Biology and Medical Genetics, University of Wisconsin–Madison

DAVID KORN [NAM], Consultant in Pathology, Massachusetts General Hospital; Professor of Pathology, Harvard Medical School

ALAN I. LESHNER [NAM], CEO Emeritus, American Association for the Advancement of Science

MARCIA McNUTT [NAS] (*ex-officio*), President, National Academy of Sciences

RICHARD A. MESERVE [NAE], Senior of Counsel, Covington & Burling LLP

C. D. MOTE, JR. [NAE] (*ex-officio*), President, National Academy of Engineering

J. SANFORD SCHWARTZ [NAM], Leon Hess Professor of Medicine, Professor of Health Care Management and Economics, School of Medicine and the Wharton School, University of Pennsylvania

CHRISTOPHER A. SIMS [NAS], Professor of Economics and Banking, Department of Economics, Princeton University

ROBERT F. SPROULL [NAE], Vice President and Director, Oracle Labs (retired); Adjunct Professor of Computer Science, University of Massachusetts at Amherst

JAMES M. TIEN [NAE], Distinguished Professor and Dean Emeritus, College of Engineering Electrical & Computer Engineering Department, University of Miami

MICHAEL S. WITHERALL [NAS], Laboratory Director, Lawrence Berkeley National Laboratory

SUSAN M. WOLF [NAM], McKnight Presidential Professor of Law, Medicine and Public Policy; Faegre Baker Daniels Professor of Law; Professor of Medicine, University of Minnesota
PETER WOLYNES [NAS], D. R. Bullard-Welch Foundation Professor of Chemistry, Center for Theoretical Biological Physics-BCR, Rice University

**Staff**

ANNE-MARIE MAZZA, Director
KEVIN FINNERAN, Director (until December 2016)
GURU MADHAVAN, Program Officer (until November 2010)
MARIA LUND DAHLBERG, Associate Program Officer (until September 2016)
JAMIE BIGLOW, Senior Program Assistant (until November 2015)
NEERAJ GORKHALY, Research Associate (until March 2014)
MARION RAMSEY, Administrative Associate (until October 2013)

# Preface

Several decades ago, a series of highly visible cases of alleged research misconduct prompted researchers, research institutions, research sponsors, and others to consider how they might promote research integrity and address breaches in integrity more effectively. Up to that time, the research enterprise and its key stakeholders often approached these issues in an informal and ad hoc manner. Ultimately, the United States and some other countries developed and implemented formal policies and procedures aimed at ensuring that research misconduct allegations are investigated, and launched new educational programs in the responsible conduct of research. The National Academies of Sciences, Engineering, and Medicine were very involved in these discussions and debates, and have made significant contributions since that time.

In recent years, as ongoing globalization, technological advances, and other shifts have transformed research, it is clear that the research enterprise faces new and complex challenges in fostering integrity and in dealing with the consequences of research misconduct and detrimental research practices. Serious cases of research misconduct—including some that have gone undetected for years—continue to emerge with disturbing regularity in the United States and around the world. Increases in the number and percentage of research articles that are retracted and growing concern about low rates of reproducibility in some research fields raise questions about how the research enterprise can better ensure that investments in research produce reliable knowledge.

It is necessary for all of us involved in performing, managing, funding, and communicating research to commit to improving practices in our own organizations and disciplines as well as more broadly. Key areas of focus include institutional efforts to sustain research environments conducive to integrity, greatly

expanded sharing of data and code, more complete reporting of results, more responsible approaches to scholarly publishing, better understanding of the causes and consequences of breaches in integrity, and clearer standards for authorship. While this report provides a framework and rationale for positive change, collective action on the part of the community will be necessary to push forward toward a research future characterized by greater integrity and quality.

I am very grateful to the committee for dedicating considerable time and effort to a project that turned out to be more difficult and time-consuming than anticipated. The experts who shared their knowledge and experience with us made a central contribution to our effort. We are also grateful to the project's sponsors, who recognized the importance of these issues. Finally, the staff members of the National Academies of Sciences, Engineering, and Medicine who worked with us were essential to performing this study, particularly Tom Arrison who has been the heart and soul of the project.

Robert M. Nerem, *Chair*

# Acknowledgments

This Consensus Study Report was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academies of Sciences, Engineering, and Medicine in making each published report as sound as possible and to ensure that it meets the institutional standards for quality, objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following individuals for their review of this report: Bruce Alberts, University of California, San Francisco; David Allison, University of Alabama, Birmingham; Stephanie Bird, *Science and Engineering Ethics*; Clyde Briant, Brown University; Philip Campbell, *Nature*; Claude Canizares, Massachusetts Institute of Technology; Arturo Casadevall, Johns Hopkins University; Peggy Fischer, National Science Foundation (retired); John Galland, University of Hawaii; Cato Laurencin, University of Connecticut; Willem Levelt, Max Planck Institute for Psycholinguistics; Melanie Loots, University of Illinois; Marcia McNutt, American Association for the Advancement of Science;[1] Barbara Redman, New York University; Dorothy Robinson, Yale University; Norman Sleep, Stanford University; Elizabeth Wager, University of Split, Croatia; and Joanne Waldstreicher, Johnson & Johnson.

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommenda-

[1] Dr. McNutt was at American Association for the Advancement of Science when she served as a reviewer.

tions of this report nor did they see the final draft before its release. The review of this report was overseen by **Georges C. Benjamin** (American Public Health Association) and **David Korn** (Massachusetts General Hospital and Harvard University). They were responsible for making certain that an independent examination of this report was carried out in accordance with the standards of the National Academies and that all review comments were carefully considered. Responsibility for the final content rests entirely with the authoring committee and the National Academies.

# Contents

**PART TWO:**
**RESEARCH MISCONDUCT AND DETRIMENTAL RESEARCH PRACTICES**

**PART THREE:**
**FOSTERING INTEGRITY IN RESEARCH**

# Summary

The scientific research enterprise is a cornerstone of modern society. In the United States alone, the public and private sectors invest hundreds of billions of dollars and countless hours of highly skilled labor into the generation, validation, and dissemination of new knowledge every year. This investment delivers enormous benefits to society in the form of better health, enhanced understanding of the natural world, and new technologies that boost economic growth and improve life in myriad ways.

The integrity of knowledge that emerges from research is based on individual and collective adherence to core values of objectivity, honesty, openness, fairness, accountability, and stewardship. When researchers commit research misconduct or engage in other behavior that clearly damages research—what this report terms detrimental research practices—they stray from the norms and appropriate practices of science. Yet the research process itself, including its system of incentives, goes beyond the actions of individual researchers. Integrity in research means that the organizations in which research is conducted encourage those involved to exemplify these values in every step of the research process: planning, proposing, performing, and reporting their work; reviewing proposals and work by others; training the next generation of researchers; and maintaining effective stewardship of the scholarly record.

The research enterprise is a complex system that includes universities and other research institutions that educate, employ, and train researchers; the federal, foundation, and industrial sponsors of research; science, engineering, technology, and medical journal and book publishers; and scientific societies. These organizations can act in ways that either support or undermine the integrity of research.

For example, research institutions may—or may not—create and maintain

research environments that support integrity, including the policies and capabilities needed to respond responsibly to allegations of research misconduct. Science, engineering, technology, and medical journal and book publishers may provide high levels of rigor in review of manuscripts, or they may put pressure on prospective authors to add citations to manuscripts to improve a journal's score on a bibliometric indicator. Fields and disciplines may take on as a community the task of defining and upholding necessary standards in areas such as data sharing, or they may fail to do so and, in effect, tolerate detrimental research practices.

Evidence accumulated over the past several decades, and particularly the past several years, provides strong support for the proposition that failing to define and respond strongly to research misconduct and detrimental research practices constitutes a significant threat to the research enterprise. This evidence is discussed mainly in Chapter 5 and Chapter 7. Highly visible research misconduct cases continue to appear regularly around the world. Appendix D describes several case studies of the multilayered challenges facing the U.S. research enterprise in fostering research integrity. Addressing threats to this integrity requires a contemporary understanding of the research system and challenges to the integrity of that system.

Concerns about scientific research that have emerged in the scientific and general media over the past several years reinforce the need to rethink and reconsider the strategies used to support integrity in research environments, including those used to prevent and address research misconduct and detrimental research practices. A growing body of evidence indicates that substantial percentages of published results in some fields are not reproducible; this lack of reproducibility appears to have many causes, ranging from essential aspects of the research process or differences in procedures to research misconduct or detrimental research practices. There also has been a remarkable increase in the number of retractions of journal articles, with analyses showing that a significant percentage of these retractions are due to research misconduct (Fang et al., 2012; Grieneisen and Zhang, 2012; Steen et al., 2013). The increase in retractions does not necessarily indicate that the incidence of misconduct is also increasing; other factors such as more vigilant scrutiny by the community and retractions becoming a more common practice among journals may be contributing factors. New forms of detrimental research practices are also appearing, such as "predatory" journals that do little or no editorial review or quality control of papers while also charging authors substantial fees and predatory conferences that charge researchers to speak at conferences that subsequently are canceled.

The research environment continues to change in significant ways that affect efforts to foster research integrity. Longstanding trends include growth in the size and scope of the research enterprise, the expansion of regulatory requirements, and an increased emphasis on industry sponsorship and entrepreneurial research. In addition, several important newer trends have emerged, including the pervasive and growing importance of information technology in research, the global-

ization of research, the increasing relevance of knowledge generated in certain fields to policy issues and political debates, and a pervasive media environment that can help generate and spread findings and controversies. These changes have led to important shifts in the institutions that support and underlie the research enterprise, such as science, engineering, technology, and medical publishing.

In assessing the trends and phenomena discussed above, along with possible new approaches, this report does not conclude that the research enterprise is broken. However, the research enterprise faces serious challenges in creating the appropriate conditions to foster and sustain the highest standards of integrity. To meet these challenges, deliberate steps must be taken to strengthen the self-correcting mechanisms that are an implicit part of research. The recommendations presented below are intended as a start to this process.

## BACKGROUND OF THE STUDY

Several decades ago, prompted by a series of high-profile cases where data fabrication, falsification, and plagiarism were alleged and investigated, the U.S. research enterprise began to institute new approaches aimed at strengthening the capacity of researchers and research institutions to foster integrity and to address research misconduct. These approaches included the development of training materials and programs in the responsible conduct of research and formal federal oversight of research misconduct investigations affecting federally funded work.

As part of these efforts, the Committee on Science, Engineering, and Public Policy (COSEPUP) of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine formed a panel to undertake a major study of issues related to scientific responsibility and the conduct of research. Completed in 1992, *Responsible Science: Ensuring the Integrity of the Research Process* recommended steps for reinforcing responsible research practices (NAS-NAE-IOM, 1992). The report also developed a framework to distinguish three categories of behaviors that can compromise the integrity of the research process. Misconduct in science was defined as "fabrication, falsification, or plagiarism in proposing, performing, or reporting research." Questionable research practices were defined as "actions that violate traditional values of the research enterprise and that may be detrimental to the research process." Other misconduct was defined as "forms of unacceptable behavior that are clearly not unique to the conduct of science, although they may occur in the laboratory or research environment." The unified federal policy adopted in 2000 uses a definition of research misconduct that largely reflects the recommendations of the COSEPUP panel.

Several years ago, COSEPUP commissioned a new committee to prepare a second edition of *Responsible Science*. In undertaking this effort, it became clear to the committee that changes in the research environment and the extent of the current challenges posed by research misconduct and other detrimental research

practices that clearly damage research required the development of a substantially new report.

## RECOMMENDATIONS

While reaffirming the central recommendation from *Responsible Science* that formally places the primary responsibility for strengthening the responsible conduct of research on individual researchers and research institutions, the committee also believes that the integrity of research depends on creating and maintaining a system and environment for research in which institutional arrangements, practices, policies, educational programs, and incentive structures support responsible conduct.

The committee also endorses the definition of research misconduct recommended in *Responsible Science*, while recommending refinements and harmonization of the definition and its use. The committee believes that many of the practices that have been categorized up to now as questionable research practices, such as the misleading use of statistics that falls short of falsification and failure to retain sufficient research data, should be recognized as detrimental to research. Detrimental research practices also should be understood to include irresponsible or abusive actions taken by research institutions and journals in addition to the actions of individual researchers.

**RECOMMENDATION ONE: In order to better align the realities of research with its values and ideals, all stakeholders in the research enterprise—researchers, research institutions, research sponsors, journals, and societies—should significantly improve and update their practices and policies to respond to the threats to research integrity identified in this report.**

Lack of attention to or tolerance of detrimental research practices by stakeholders makes it difficult to expose misconduct, wastes human and financial resources, impairs the overall quality of research, and diminishes public trust in science. In addition, weaknesses in the system for identifying, investigating, and addressing research misconduct—most notably unevenness in the policies and capabilities of research institutions and science, engineering, technology, and medical journal and book publishers—create barriers to uncovering misconduct and taking appropriate action. Similarly, in industry-performed or industry-sponsored research, pressures associated with regulatory approvals or commercial release may create disincentives for full data transparency or biases that favor conclusions of safety and efficacy. Finally, changes in the research environment such as technological advances and globalization are making it more difficult and complex for all stakeholders in the enterprise to update and ensure adherence to best practices.

The checklists presented in Chapter 9 should form the basis of strategies to refine and implement best practices by researchers, research institutions, research sponsors, journals, and societies.

**RECOMMENDATION TWO: Since research institutions play a central role in fostering research integrity and addressing current threats, they should maintain the highest standards for research conduct, going beyond simple compliance with federal regulations in undertaking research misconduct investigations and in other areas.**

The key responsibilities for research institutions fall into four areas. The first is creating and sustaining a research culture that fosters integrity and encourages adherence to best practices. This includes maintaining education and training efforts that support a culture of integrity, consistent with the current state of knowledge (see Recommendation Ten).

A second task is monitoring the integrity of research environments. Such monitoring is critical to further advance understanding of how institutional structure, context, and incentives interact to buttress or detract from research integrity. Research organizations have an obligation to assess, monitor, and work to implement improvements to their research environments. Where institution-wide assessments identify units with particularly strong integrity environments, they should be examined and their practices should be disseminated and emulated.

The third institutional responsibility is ensuring that research institutions sustain the capacity needed to effectively investigate and address allegations of research misconduct. No institution can be expected to prevent all lapses in research integrity, but all should ensure that when problems in the conduct of research are alleged or identified there is a prompt, effective, and documented response to the allegation.

A fourth responsibility is ensuring that senior institutional leaders, including the president, other senior executives, and faculty leaders, are guiding and actively engaged in the preceding three tasks. When institutional leaders are accessible and knowledgeable about institutional capacity to address allegations of misconduct, they are in a position to help keep people and processes on track when specific allegations arise. Should later events call into question the rigor of an institutional response to allegations of misconduct in research, top institutional leadership should be expected, as a matter of course, to examine the shortcomings of the process and share lessons learned with the larger community of scholars. Institutional leaders should also reiterate the importance of critical standards such as appropriate authorship practices, data sharing, and complete reporting of results.

**RECOMMENDATION THREE: Research institutions and federal agencies should work to ensure that good-faith whistleblowers are**

**protected and that their concerns are assessed and addressed in a fair, thorough, and timely manner.**

Those who raise concerns about the integrity of research, often referred to as whistleblowers, can play a critical role in supporting best practices in research and in uncovering research misconduct. Individuals closest to the research are in the best position to identify and correct problems as early as possible and can be expected to play this role for the foreseeable future. Inadequate responses to expressed concerns have constituted a critical point of failure in many cases of misconduct where investigations were delayed or sidetracked. Those who raise concerns are often the most vulnerable participants in the system, typically holding little institutional power or status. Research institutions and federal agencies should understand the implicit bias that exists against those who in good faith raise fact-based concerns about the integrity of research.

**RECOMMENDATION FOUR: To provide a continuing organizational focus for fostering research integrity that cuts across disciplines and sectors, a Research Integrity Advisory Board (RIAB) should be established as an independent nonprofit organization. The RIAB will work with all stakeholders in the research enterprise—researchers, research institutions, research sponsors and regulators, journals, and scientific societies—to share expertise and approaches for addressing and minimizing research misconduct and detrimental research practices. The RIAB will also foster research integrity by stimulating efforts to assess research environments and to improve practices and standards.**

While various groups, institutions, and individuals are doing valuable work to foster and promote research integrity in the United States, no permanent organizational focus for efforts to foster research integrity at a national level currently exists. The Research Integrity Advisory Board recommended by the committee would bring a unified focus to understanding and addressing challenges across all disciplines and sectors.

The RIAB could facilitate the exchange of information regarding approaches to assessing and creating environments of the highest integrity and to the handling of allegations of misconduct and investigations. It could provide advice, support, encouragement, and, where helpful, advocacy on what needs to be done by research institutions, science, engineering, technology, and medical journal and book publishers, and other stakeholders in the research enterprise to promote research integrity. The RIAB will have no direct role in investigations, regulation, or accreditation. Rather, it will serve as a neutral resource based in the research enterprise that helps the enterprise respond to ongoing and future changes.

**RECOMMENDATION FIVE: Societies and journals should develop clear disciplinary authorship standards. Standards should be based on the principle that those who have made a significant intellectual contribution are authors. Significant intellectual contributions can be made in the design or conceptualization of a study, the conduct of research, the analysis or interpretation of data, or the drafting or revising of a manuscript for intellectual content. Those who engage in these activities should be designated as authors of the reported work, and all authors should approve the final manuscript. In addition to specifying all authors, standards should (1) provide for the identification of one or more authors who assume responsibility for the entire work, (2) require disclosure of all author roles and contributions, and (3) specify that gift or honorary authorship, coercive authorship, ghost authorship, and omitting authors who have met the articulated standards are always unacceptable. Societies and journals should work expeditiously to develop such standards in disciplines that do not already have them.**

Authorship practices are a fundamental component of the research enterprise's operation, and observance of good practices is a key factor in ensuring research integrity. Authorship crucially designates who bears responsibility for the work. Clarifying authorship responsibility is also critical in cases of error or allegations of misconduct.

The current situation, in which authorship practices and conventions are largely left to individual institutions and journals, is increasingly problematic. Greater clarity at the disciplinary level about the significant intellectual contributions that merit authorship, the roles that do not merit authorship, the significance of author order, and the responsibilities of a primary or corresponding author would be very helpful in facilitating appropriate decisions and practices in laboratories and collaborations. Universal condemnation (i.e., by all disciplines) of gift or honorary authorship, coercive authorship, and ghost authorship would also contribute to changing the culture of research environments where these practices are still accepted.

**RECOMMENDATION SIX: Through their policies and through the development of supporting infrastructure, research sponsors and science, engineering, technology, and medical journal and book publishers should ensure that information sufficient for a person knowledgeable about the field and its techniques to reproduce reported results is made available at the time of publication or as soon as possible after that.**

In many fields and disciplines, current standards for transparency are not adequately supporting reproducibility and the ability to build on previous work. However, the research enterprise has begun to take important steps. Some journals have begun to implement requirements that authors make the data and computer code required to regenerate the published results available upon request. Many universities and funding agencies have created online repositories to support the dissemination of digital data. Current data practices vary significantly by field and discipline, and making certain types of data broadly accessible presents special challenges. The successful development and implementation of new standards and requirements will depend upon sufficient investments in necessary human and physical infrastructure.

**RECOMMENDATION SEVEN: Federal funding agencies and other research sponsors should allocate sufficient funds to enable the long-term storage, archiving, and access of datasets and code necessary for the replication of published findings.**

Preparing data and code for release can be expensive and time consuming. Researchers are currently rewarded for manuscript publication, but the professional rewards for preparing data and code for publication are minimal. The resources to support the endeavor are often limited, and the feasibility and time required depend on the type of research data and how those data were collected.

Journals should update their publication requirements to include access to data and code needed to replicate results in the manuscript. These data and codes can be deposited at any repository that can reasonably guarantee a persistent URL, to be provided in the text of the published paper. To facilitate the reuse of scientific code and data, these objects should be shared in a way that maximizes access while respecting scientific norms such as attribution.

**RECOMMENDATION EIGHT: To avoid unproductive duplication of research and to permit effective judgments on the statistical significance of findings, researchers should routinely disclose all statistical tests carried out, including negative findings. Research sponsors, research institutions, and journals should support and encourage this level of transparency.**

Today, several initiatives exist to encourage and promote reproducibility in research. As routine reporting of negative results and statistical tests becomes the standard for all fields, research spending will become more productive, with more knowledge generated per dollar of research investment. Changing the culture of research and publication so that reporting negative results is required will depend on a persistent effort on the part of disciplines, sponsors, and journals.

**RECOMMENDATION NINE: Government agencies and private foundations that support science, engineering, and medical research in the United States should fund research to quantify, and develop responses to, conditions in the research environment that may be linked to research misconduct and detrimental research practices. These research sponsors should use the data accumulated to monitor and modify existing policies and regulations.**

While understanding of the causes and incidence of research misconduct and detrimental research practices has increased, critical knowledge gaps remain. For example, official statistics on findings of research misconduct may represent a lower bound on incidence, with survey data pointing to a significantly higher incidence of misconduct, but no reliable estimate of incidence or trends exists. Also, detrimental research practices are more widespread and may ultimately be more damaging to the research enterprise than research misconduct, which points to the need to address challenges to research integrity more broadly. In addition, trends in some indicators—such as declining success rates for grant applications, and an increasing ratio of PhD production to available faculty positions—raise the possibility that both local organizational environments and the broader structural arrangements of research are moving in directions that might threaten research integrity. Additional theoretically grounded research with subsequent testing in practice is warranted to more completely inform efforts to improve research environments and incentive structures.

**RECOMMENDATION TEN: Researchers, research sponsors, and research institutions should continue to develop and assess more effective education and other programs that support the integrity of research. These improved programs should be widely adopted across disciplines and across national borders.**

Formal education and training in the responsible conduct of research (RCR) can play an important role in fostering integrity and strengthening research environments. Evidence developed to date indicates that much remains to be learned about the approaches that are most effective. RCR education can serve as a key element in strategies to promote integrity, but perhaps not as the primary means of addressing research misconduct and detrimental research practices in the short term. Evidence-based assessment and improvement of RCR education programs are needed, with the focus expanded to include the social and institutional environment for research. RCR education should engage not only junior researchers but also senior researchers and industrial researchers.

**RECOMMENDATION ELEVEN: Researchers, research institutions, and research sponsors that participate in and support international collaborations should leverage these partnerships to foster research integrity through mutual learning and sharing of best practices, including collaborative international research on research integrity.**

While the committee was tasked with considering the issue of research integrity regarding U.S.-based institutions and U.S. policies, a good deal of research now takes place in other countries and across national boundaries. Given that research misconduct, detrimental research practices, and the need to foster research integrity are challenges facing all countries that fund and perform research, the global research enterprise will benefit from the knowledge gained from examining research practices globally.

# Part One

# The Integrity of Research

# 1

# Introduction

*The public will support science only if it can trust the scientists and institutions that conduct research.*

—National Research Council and Institute of Medicine (2002)

The achievements of science in formulating a systematic knowledge of the physical, biological, and social world have been breathtaking. Study of both the very large and the very small has revealed that the universe began with an initial singularity and has deepened our knowledge of its expansion over more than 13 billion years. Research into the fundamental building blocks that make up living things has unlocked mysteries of heredity, biochemical bases of diseases, and pathways to improved medicines and better health. Examination of brain function is providing new insights into learning, emotion, and mental disorders. Studies of human communities have contributed knowledge of psychological, social, and political behaviors to inform a continued expansion of human well-being consistent with environmental sustainability.

This astonishing growth in human knowledge has been accompanied by a dramatic explosion of technologies designed to meet human needs and enhance human well-being. New drugs and devices, including imaging devices based on research into the properties of materials, have contributed to the extension and improvement of human life. The development of digital technologies has not only expanded human capabilities but has created entirely new ways of communicating, sensing, analyzing, learning, and doing research. Advances in agriculture, transportation, and food science have increased human capacities to feed a growing world population. Often drawing inspiration from scientific research and sometimes enabling that research, technological and other forms of innovation have become a mainstay of modern economies.

A major contributor to the remarkable progress of science and technology has been the ability of the research enterprise to continuously build upon a foundation of knowledge created by previous researchers. The stability of this founda-

tion has resulted from the adherence of researchers to good research practices and from their openness in communicating research procedures and results. By communicating the assumptions made and methods used in conducting experiments, researchers allow others to replicate, extend, and, where necessary, correct their work. When undertaken with fidelity, science becomes a cumulative exercise that produces a growing body of reliable knowledge.

Science progresses through both success and failure. Modest gains are not just incremental but are interspersed in unpredictable ways with huge breakthroughs followed by periods of consolidation. Research frequently returns negative results, and such failures are necessary to push the boundaries of knowledge forward. Even when researchers are careful and scrupulous, results will be reported that later turn out to be the result of incomplete understanding, honest errors in performing experiments and interpreting data, or limitations in the capabilities of research instrumentation available at a particular time. Adherence to good practices and values such as openness and transparency minimizes missteps and increases the likelihood of efficient progress toward new, reliable knowledge: it enables good science.

The research enterprise is a system of individuals, organizations, and relationships that requires its constituents to fulfill their responsibilities in order to be effective. In contrast to simple systems, which are stable and whose components interact through well-understood cause-and-effect relationships, the research enterprise is more akin to a complex adaptive system characterized by dynamism and self-organization. Such systems "can be highly organized without any conscious leadership, direction, or management" and exist "within other interdependent systems" (McKenzie, 2014). The components of complex adaptive systems change as they interact with each other. Cause-and-effect relationships within the system are influenced by feedback effects, and "changes in one part of the system can cause changes in other parts of the system, often in nonlinear and unpredictable ways" (McKenzie, 2014).

Figure 1-1 illustrates the research enterprise as a complex adaptive system. The figure, which is meant to be stylized and heuristic rather than purely and exhaustively descriptive, distinguishes two conceptually distinct "components" of the system: participants and stakeholders, and systems and processes. These components interact with one another and across the two types in ways that may be bidirectional, implying the existence of nonlinearity, feedback loops, and complex causality. As discussed in other parts of this report, some components of the research enterprise operate at a local level (research institutions) or a national level (research funding systems, to a large extent). Other components, such as publication and the dissemination of knowledge, operate largely at a global level.

Some researchers deviate from the norms and practices that they are expected to fulfill. The reasons can be complex, including intentional or negligent actions resulting from carelessness or other individual shortcomings coupled with environmental pressures and institutional practices. Deviations from good science

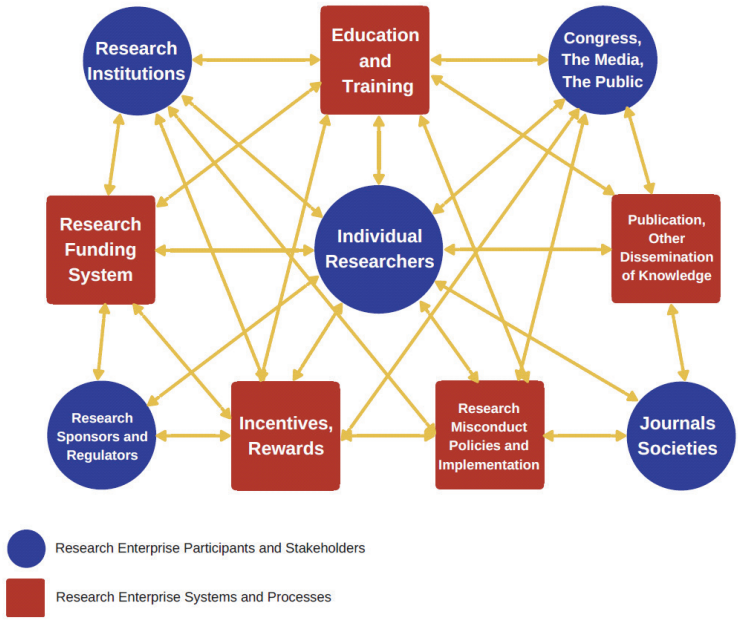


FIGURE 1-1 The research enterprise is a complex adaptive system.

can cause great damage to the research enterprise—both in the practice of science and in how that science is perceived in the broader society. Organizations such as research institutions, research sponsors, and journals may also engage in detrimental research practices. Even fields and disciplines may fail to define and uphold necessary standards in areas such as data sharing, in effect tolerating detrimental research practices.

Making matters more complex, in recent years the research enterprise has been changing in ways that can make both the identification and the application of best practices more difficult than in the past. For example, ensuring research integrity may require encouraging and rewarding behaviors that have not been valued in the past, such as publication of negative results.

A central goal of this report is to identify best practices in research and to recommend practical options for discouraging and addressing research misconduct and detrimental research practices. The sustainability of the scientific enterprise, both as a body of practice and as a legitimate, authoritative contributor to societal ends, depends in no small part on putting best practices to work across the entire system.

## THE 1992 *RESPONSIBLE SCIENCE* REPORT

In 1989, the Committee on Science, Engineering, and Public Policy of the National Academy of Sciences, National Academy of Engineering, and Institute of Medicine formed a 22-member panel to conduct a major study of issues related to scientific responsibility and the conduct of research. The goals of the panel were to review factors affecting the integrity of science and the research process, to recommend steps for reinforcing responsible research practices, to review institutional mechanisms for addressing allegations of misconduct in science, and to consider the advantages and disadvantages of formal guidelines and enhanced educational efforts for the conduct of research.

In 1992, the panel released its report *Responsible Science: Ensuring the Integrity of the Research Process* (NAS-NAE-IOM, 1992). The panel defined the term "integrity of the research process" as "the adherence by scientists and their institutions to honest and verifiable methods in proposing, performing, evaluating, and reporting research activities." The panel also developed a framework that distinguishes three categories of behaviors that can compromise the integrity of research. *Misconduct in science* was defined as "fabrication, falsification, or plagiarism in proposing, performing, or reporting research." *Questionable research practices* were defined as "actions that violate traditional values of the research enterprise and that may be detrimental to the research process." *Other misconduct* was defined as "forms of unacceptable behavior that are clearly not unique to the conduct of science, although they may occur in the laboratory or research environment."

## WHY REVISIT THESE ISSUES?

The 1992 *Responsible Science* report provided a valuable service in describing and analyzing a very complicated set of issues, and it has served as a crucial basis for thinking about research integrity for more than two decades. However, as experience has accumulated with various forms of research misconduct, detrimental research practices, and other forms of misconduct, as subsequent empirical research has revealed more about the nature of scientific misconduct, and because technological and social changes have altered the environment in which research is conducted, it is clear that the framework established more than two decades ago needs to be updated. In order to develop more robust approaches to ensuring research integrity in the current research environment, it is necessary to revisit some of the issues addressed in *Responsible Science*.

A recent case illustrates some of the questions and issues that should be rethought. On August 5, 2014, stem cell biologist Yoshiki Sasai of Japan's RIKEN (a large research institution) was found dead in an apparent suicide. Sasai had been the supervisor of Haruko Obokata, another RIKEN researcher, and one of Obokata's coauthors on two papers published in the journal *Nature* purporting to have discovered an easy way to transform ordinary cells into pluripotent stem

cells, which would give a significant boost to a variety of therapies that utilize stem cells (Obokata et al., 2014a,b). Earlier versions of the work had been rejected by *Nature* and several other journals before being published in early 2014 (Vogel and Normile, 2014). Questions about the work arose almost immediately, as many researchers around the world set out to replicate and extend the remarkable results and failed. RIKEN promptly conducted an investigation, concluding that Obokata had fabricated and falsified data (Ishii et al., 2014) (Appendix D provides a full case description).

This case shares many elements with several other high-profile misconduct cases that have emerged in recent years, including (1) a striking, counterintuitive result trumpeted in a prestigious journal; (2) the apparent failure of supervisors at the lab and institutional levels as well as other coauthors to check or question data that turned out to be fairly obviously fabricated or falsified; (3) international coauthorship that illustrates the global nature of challenges to integrity; and (4) a publication whose results are immediately cast into doubt by the relevant research community, perhaps because the publication is in a high-profile field where efforts to replicate and extend the work would be expected to commence soon after publication. Given that cases with similar elements appear fairly regularly, it is fair to ask several questions: *Could* better research practices at the lab or institutional levels, at journals, and at the field or disciplinary level prevent a significant fraction of these cases? If so, can better practices be developed and implemented so that the behavior of researchers actually changes?

The need for rethinking and reconsideration of approaches to understanding, preventing, and addressing research misconduct and detrimental research practices is reinforced by a narrative that has emerged in the scientific and general media over the past several years: that the research enterprise itself is somehow broken or seriously off track (Fang et al., 2012; *Economist*, 2013; Alberts et al., 2014). Some of the phenomena that have driven this narrative, in addition to the regular appearance of highly visible research misconduct cases, include growing evidence that half or more of the published results in some fields are not reproducible, the remarkable growth in the number of retractions of journal articles, and the appearance of new forms of detrimental research practices such as journals that charge authors to publish but appear to do no quality control. According to the *Economist*, "Fraud is very likely second to incompetence in generating erroneous results, though it is hard to tell for certain" (*Economist*, 2013).

The trends and phenomena listed above are discussed in detail in the report, along with possible new approaches. For the purpose of this introduction, it is important to note that the report does not conclude that science itself is broken. Rather, the research enterprise faces serious challenges in dealing with research misconduct and detrimental research practices in the current environment. For example, detrimental research practices are more widespread and may ultimately be more damaging to research than research misconduct is. In order to meet these challenges and secure the future, the research enterprise needs to make deliberate

efforts to strengthen the self-correcting mechanisms that are already an implicit part of science.

The report also concludes that, while the core values of research do not and should not change, there is a need to restate and reconfirm these values for the 21st century. Research is based on a set of enduring principles that have proven to be successful in generating empirically based knowledge of the natural world. It also takes place using particular practices, techniques, and methods that vary from one research field to another, across research groups, and over time. The ways in which science is carried out and communicated have evolved dramatically over the past two decades. The research enterprise in the United States and around the world has undergone tremendous change over this time, and these changes can have implications for how the enterprise fosters integrity. Research takes place within particular contexts that can have a powerful influence on the productivity and applicability of research both within science and in the broader society. These methods and contexts have been changing rapidly over the past two decades; this also has created new challenges to upholding research integrity.

## A CHANGING ENVIRONMENT

*Responsible Science* contains a chapter entitled "Contemporary Research Environment" that laid out the changes in science from the post–World War II period to the early 1990s. It noted that in the 1960s a typical laboratory research group might have consisted of less than a half-dozen members, while in the 1990s larger and more diverse groups were becoming more common. It cited "the increasing complexity of contemporary research problems and instrumentation, the increasing costs of scientific research, changes in the rationale for supporting and monitoring government-funded research, and increased regulation of federal research." As the number of government regulations and guidelines had increased, the report observed, universities had expanded administrative and oversight functions, which had the potential to create or exacerbate tensions between administrators and faculties. Also, the growing interest of private industry in research results was leading to more cooperative programs between universities and industry, encouraged by federal and state programs designed to foster such cooperation.

All of these trends have intensified since the early 1990s. Today, the research system is even larger, more complex, and more integrated into other societal sectors. In the United States, the number of science, engineering, and health doctorate holders employed in academia rose nearly 30 percent, from 211,000 in 1991 to almost 309,000 in 2013 (NSB, 2016). The number of PhDs awarded in science and engineering rose 95 percent, from approximately 19,000 in 1988 to almost 37,000 in 2013, with an increasing percentage of these doctorate recipients going to work outside academia (NSB, 2016). The number of science and engineering articles worldwide grew more than 350 percent, from approximately 485,000 in

1989 to approximately 2,200,000 in 2013, with especially rapid growth outside the United States (NSB, 2016). Government regulation of research has continued to grow, to the point where its "ever-growing requirements are diminishing the effectiveness of the nation's research investment" (National Academies of Sciences, Engineering, and Medicine, 2016).

Even as expansion and diversification have continued, new developments have reshaped the research enterprise since the early 1990s. These include:

*Collaboration*. The increasing prevalence of multi-investigator, interdisciplinary research has led to the creation of research teams where sometimes no one member can understand the details of all of the science encompassed by the research. In such circumstances, collaborative research must include structures that coordinate and verify the integrity of separate contributions to the overall research effort. These efforts are further complicated by the fact that collaborations often operate across institutional and national boundaries (see below).

*Globalization*. While research has always been international in many important respects, the scale and scope of research practice are globalizing to an unprecedented extent. This is seen in the spread of capabilities around the world (particularly in large emerging economies such as China, India, South Korea, and Brazil), the growth of large- and small-scale collaborations across borders, and continued internationalization of the U.S. research workforce (described in more detail in Chapter 3). Differences in culture, language, and training can lead to lack of understanding and clarity about values, roles, and responsibilities and could contribute to research environments where it is more difficult to prevent, uncover, or respond to lapses in integrity.

*Funding and Organization*. The funding and organization of research in the United States and the institutions that perform and communicate it have undergone some changes. For example, the share of U.S. research and development funded by industry has grown somewhat (from 58 percent in 1992 to 65 percent in 2013) while the share funded by government has shrunk (from 36 percent in 1992 to 27 percent in 2013) (NSB, 2016). However, looking only at basic and applied research and leaving development aside, the federal share of funding has held steady at about 45–46 percent while the industry share has declined from 44 percent to 36 percent. The commercialization of university research has increased over the years, as measured by patenting, licensing, and the launch of start-up companies based on university technology (AUTM, 2014).

Some of the key institutions for performing and communicating research, including research universities and scholarly journals, are experiencing significant stresses. The organizational model for academic research is shifting, with the results perhaps most visible in biomedicine: larger research groups, lower success rates on grants, a growing population of graduate students and postdoctoral fellows who are less likely to attain tenure-track or other independent research positions than previous generations, and the greater reliance of researcher sala-

ries on federal support that has sometimes moved in unpredictable ways. U.S. research universities and higher education institutions more broadly face a range of long-term pressures apart from shifts in research funding, including pressures on general funding from states (for public universities), tuition growth that has outpaced increases in the overall cost of living, demographic shifts, and disruptions caused by the advent of new technologies (NRC, 2012a).

*Competitive Pressure*. Increasing pressure on both junior and senior researchers to publish in prominent journals has created a bias to produce the kinds of novel, newsworthy, and paradigm-shifting results favored by these journals. Similarly, the difficulty in securing government grants and contracts along with explicit federal requirements to do so have increased the pressure on researchers to emphasize the significance and relevance of proposed research. The importance of publications in establishing the reputation of researchers and as the basis for hiring and promotion decisions has increased the potential for disputes over authorship and distorts the publication process—for example, by heightening the temptation to publish multiple papers on just one experiment or dataset.

*Policy and Societal Relevance*. The relationship between the research enterprise and the larger society, including policy makers and the public, has become deeper and more complex. Research is implicated in more policy areas with higher stakes, so as science is called upon to inform decision making there is more risk of research being invoked in controversies, misrepresented, or shaped to advance a desired political outcome, contributing to poor decision making and loss of public trust.

*Technological Changes*. Research in most fields has been transformed by the advance of technology, particularly the emergence of approaches to research across many fields that take advantage of the ability to collect and analyze large amounts of digital data and the infusion of information technologies into communications. As will be explored further in the report, technological advances have enabled new ways for researchers to err—both intentionally and unintentionally—as well as offering new tools to detect mistakes and misbehavior.

*Scholarly Communications*. New forms of scientific publication pose challenges to traditional peer review systems.[1] Examples include non-peer-reviewed web publications that are widely available, "publication" on personal web pages, and rapid publication with continuously updated reviews. The emergence of research based on computer analyses of massive datasets raises questions about access to both the data and the computer code used to analyze the data and about the allocation of credit to those who collect, curate, and disseminate data and to those who create software and programs that perform scientific analysis on the datasets. Computational science also raises questions regarding appropriate stewardship and persistence of datasets and code.

[1] This report generally uses the term *peer review* to refer to systems that bring expert perspectives to bear on the evaluation of articles submitted for publication and the evaluation of grant proposals. Alternative terms include *merit review* and *expert review*.

*Visibility of Research Misconduct and Detrimental Research Practices*. Research misconduct is reported on in the science media and in some cases the general media, with a steady stream of cases from around the world. Retractions and other indicators related to research misconduct and detrimental research practices are on the rise, and there are new mechanisms for communicating cases and trends. Policy reports on research integrity are emerging from a variety of international groups and individual countries and from the large international community of scholars, educators, and other practitioners concerned with these issues. This is reflected in phenomena such as the world conferences on research integrity and the launch of the Association of Research Integrity Officers. More detailed discussion about these trends appears elsewhere in the report. The overall result is that research misconduct and detrimental research practices are becoming more visible and attracting more attention.

*Evolution of Policies*. The U.S. policy framework related to research misconduct has evolved over the past two decades. Changes such as the 2000 unified federal definition and new or revised education mandates have emerged within a framework where primary responsibility for investigating and taking corrective actions in response to misconduct and other undesirable behavior lies with institutions, overseen by funding agencies. There has also been growing focus on research integrity issues around the world, with a number of individual countries adopting or changing policies, as well as reports by international bodies such as the InterAcademy Partnership (IAC-IAP, 2012; IAP, 2016). In light of other changes in the research environment, it is worth examining the current policy framework and related administrative procedures to see how they are working and whether they are adapted to research as it is performed today and as it will likely evolve in coming decades.

*Improved Understanding of Research Misconduct and Detrimental Research Practices*. A significant change since the publication of *Responsible Science* is the accumulation of knowledge through the passage of time. More cases of research misconduct, the development of empirical research on research misconduct and detrimental research practices, and better understanding of human cognition and decision making have all contributed to this knowledge. More is known about the incidence of major and "minor" departures from research norms, and about the factors that can influence behavior inside organizations. Examining the systems within which research takes place and considering how these can be designed and managed in ways that buttress and reinforce the integrity of research provides better understanding of the emerging threats to research integrity and can lead to policies and interventions that address these threats.

Adding a focus on how research environments and systemic conditions influence individual choices does not lessen the personal responsibility of researchers for their actions. Nor does it lessen the importance of educating students and working professionals about their roles in upholding the integrity of the community of scholars. Each individual retains deeply personal obligations and duties

**BOX 1-1**
**Statement of Task**

An ad hoc committee under the oversight of the Committee on Science, Engineering, and Public Policy will undertake a revision of the Responsible Science study first issued in 1992. The committee will be charged with addressing the following questions:

- What is the state of current knowledge about modern research practices for a range of disciplines, including trends and practices that could affect the integrity of research? What is the impact of modern technology such as image enhancement, the Internet, and data storage systems?
- What are the impacts on integrity of changing trends in the dynamics of the research enterprise, such as globalization, the treatment of intellectual property, handling of materials and specimens, university oversight and institutional review boards, and demands of government regulation?
- What are the advantages and disadvantages of enhanced educational efforts and explicit guidelines for researchers and research institutions? Can the research enterprise itself define and strengthen basic standards for scientists and their institutions? How is this affected by increased collaboration among researchers, in the United States and internationally?
- What roles are appropriate for government agencies, research institutions and universities, and journals in promoting responsible research practices? What can be learned from institutional and journal experiences with current procedures for handling allegations of misconduct in science?
- What should the definition of research misconduct include? Should it only include the criteria of "falsification, fabrication, and plagiarism" (drawn from the 1992 edition of *Responsible Science*) or should it be broadened to include elements of questionable research practices and research impropriety?
- Should existing unwritten practices be expressed as principles to guide the responsible conduct of research? The committee is encouraged to prepare model guidelines and other materials if it deems that would be useful.

to aspire to the highest levels of rigor in his or her work, including understanding about cognitive biases and errors to which decision-making processes are prone in order to build in safeguards and precautions against natural influences in favor of existing ideas, personal biases, the interests of funders, and the reward systems that surround the pursuit of science.

## STRUCTURE OF THE REPORT

This report is divided into three parts. The first part focuses on integrity in research, including this introduction, the underlying values and norms of research, and shifts in the research system and how they affect integrity. The second part covers research misconduct and detrimental research practices, including definitions, a historical overview of how these issues have been handled by the research enterprise, incidence and consequences, understanding why researchers commit these transgressions, and how they are addressed by various stakeholders. The third part considers how the research enterprise can better foster integrity in research, including education and training, defining best practices and encouraging their adoption, and the report's findings and recommendations.

See Box 1-1 for the committee's task statement. The first two task elements are largely addressed in Chapter 3, with some material addressing these elements in Chapters 5, 6, and 7.[2] The third task element is largely addressed in Chapters 7, 8, 9, and 10. The fourth task element is largely addressed in Chapters 4, 7, 8, and 9. The fifth task element is largely addressed in Chapters 4 and 7. The sixth task element is largely addressed in Chapters 7, 8, and 9. All of the task elements are addressed to some extent in Chapter 11. Box 1-2 contains further information about terminology used in the study.

[2] Although this report contains some discussion of federal regulations covering the treatment of human research subjects and how this regulatory framework interacts with implementation of the federal government's research misconduct policy, the committee considered institutional review boards and other elements of human subjects protections to be outside the appropriate scope of its findings and recommendations. The federal government was in the process of developing changes to the regulatory framework designed to protect human subjects—the Common Rule—during the course of this study. The National Academies of Sciences, Engineering, and Medicine were undertaking a comprehensive study of government research regulations at the same time that this study was in process (National Academies of Sciences, Engineering, and Medicine, 2016).

**BOX 1-2**
**Definitions**

**Research:** The Merriam-Webster dictionary defines research as "studious inquiry or examination; especially: investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws." The federal research misconduct policy defines research as including:

> all basic, applied, and demonstration research in all fields of science, engineering, and mathematics. This includes, but is not limited to, research in economics, education, linguistics, medicine, psychology, social sciences, statistics, and research involving human subjects or animals. (OSTP, 2000)

In practical terms, this report is primarily directed toward the many stakeholders involved in scientific research, including the natural sciences, social and behavioral sciences, clinical medical research, and engineering research. Although research in the humanities, the law, or other endeavors is not the primary focus, some conclusions and insights in this report may be usefully applied to those fields. The terms *science* and *research* are often used interchangeably in this report.

**Research misconduct:** The 1992 report defined research misconduct as fabrication, falsification, and plagiarism (FFP), and the Office of Science and Technology Policy adopted FFP as the unified federal definition in 2000. This committee accepts and builds on this stance. As discussed in Chapter 4, although there is

broad agreement around the world that fabrication, falsification, and plagiarism are included in research misconduct, other behaviors are included in definitions used by some U.S. research institutions and by some other countries. In this report, research misconduct will mean FFP except in the context of discussing institutional and international differences. In discussion of U.S. policy, the federal research misconduct definition will be specified as such.

**Detrimental research practices:** Detrimental research practices are research practices other than FFP that are clearly detrimental to the research process, as explained more fully in Chapter 4. This report recommends that this term be used instead of *questionable research practices*.

**Other misconduct:** As described in the 1992 report, other misconduct is unacceptable behavior that is not unique to the research environment.

**Best practices in research:** As described more fully in Chapter 9, best practices in research are those behaviors undertaken by individuals and organizations that are based on the core values of science and enable good research.

**Whistleblower:** A whistleblower is someone who in good faith brings concerns about possible research misconduct or detrimental research practices to the attention of others within or outside the organization where misconduct or detrimental research practices might be occurring. Concerns can be raised informally or in formal allegations. Although the term is increasingly used in neutral or positive contexts, it unfortunately still carries a negative connotation for some. The committee was not able to identify an appropriate alternative.

# 2

# Foundations of Integrity in Research: Core Values and Guiding Norms

*Problems of scientific freedom and responsibility are not new; one need only consider, as examples, the passionate controversies that were stirred by the work of Galileo and Darwin. In our time, however, such problems have changed in character, and have become far more numerous, more urgent and more complex. Science and its applications have become entwined with the whole fabric of our lives and thoughts. . . . Scientific freedom, like academic freedom, is an acquired right, generally accepted by society as necessary for the advancement of knowledge from which society may benefit. Scientists possess no rights beyond those of other citizens except those necessary to fulfill the responsibility arising from their special knowledge, and from the insight arising from that knowledge.*

—John Edsall (1975)

***Synopsis:*** *The integrity of research is based on adherence to core values—objectivity, honesty, openness, fairness, accountability, and stewardship. These core values help to ensure that the research enterprise advances knowledge. Integrity in science means planning, proposing, performing, reporting, and reviewing research in accordance with these values. Participants in the research enterprise stray from the norms and appropriate practices of science when they commit research misconduct or other misconduct or engage in detrimental research practices.*

## TRANSMITTING VALUES AND NORMS IN RESEARCH

The core values and guiding norms of science have been studied and written about extensively, with the work of Robert Merton providing a foundation for subsequent work on the sociology of science (Merton, 1973). Merton posited a set of norms that govern good science: (1) Communalism (common ownership of scientific knowledge), (2) Universalism (all scientists can contribute to the advance of knowledge), (3) Disinterestedness (scientists should work for the good of the scientific enterprise as opposed to personal gain), and (4) Organized Skepticism (results should be examined critically before they are accepted). Research on scientists and scientific organizations has also led to a better understanding of

counternorms that appear to conflict with the dominant Mertonian norms but that are recognized as playing an inherent part in the actual practice of science, such as the personal commitment that a scientist may have to a particular hypothesis or theory (Mitroff, 1974).

More recent work on the effectiveness of responsible conduct of research education, covered in more detail in Chapter 9, explores evidence that at least some scientists may not understand and reflect upon the ethical dimensions of their work (McCormick et al., 2012). Several causes are identified, including a lack of awareness on the part of researchers of the ethical issues that can arise, confidence that they can identify and address these issues without any special training or help, or apprehension that a focus on ethical issues might hinder their progress. An additional challenge arises from the apparent gap "between the normative ideals of science and science's institutional reward system" (Devereaux, 2014). Chapter 6 covers this issue in more detail. Here, it is important to note that identifying and understanding the values and norms of science do not automatically mean that they will be followed in practice. The context in which values and norms are communicated and transmitted in the professional development of scientists is critically important.

Scientists are privileged to have careers in which they explore the frontiers of knowledge. They have greater autonomy than do many other professionals and are usually respected by other members of society. They often are able to choose the questions they want to pursue and the methods used to derive answers. They have rich networks of social relationships that, for the most part, reinforce and further their work. Whether actively involved in research or employed in some other capacity within the research enterprise, scientists are able to engage in an activity about which they are passionate: learning more about the world and how it functions.

In the United States, scientific research in academia emerged during the late 19th century as an "informal, intimate, and paternalistic endeavor" (NAS-NAE-IOM, 1992). Multipurpose universities emphasized teaching, and research was more of an avocation than a profession. Even today, being a scientist and engaging in research does not necessarily entail a career with characteristics traditionally associated with professions such as law, medicine, architecture, some subfields of engineering, and accounting. For example, working as a researcher does not involve state certification of the practitioner's expertise as a requirement to practice, nor does it generally involve direct relationships with fee-paying clients. Many professions also maintain an explicit expectation that practitioners will adhere to a distinctive ethical code (Wickenden, 1949). In contrast, scientists do not have a formal, overarching code of ethics and professional conduct.

However, the nature of professional practice even in the traditional professions continues to evolve (Evetts, 2013). Some scholars assert that the concept of professional work should include all occupations characterized by "expert knowledge, autonomy, a normative orientation grounded in community, and

high status, income, and other rewards" (Gorman and Sandefur, 2011). Scientific research certainly shares these characteristics. In this respect, efforts to formalize responsible conduct of research training in the education of researchers often have assumed that this training should be part of the *professional* development of researchers (IOM-NRC, 2002; NAS-NAE-IOM, 1992). However, the training of researchers (and research itself) has retained some "informal, intimate, and paternalistic" features. Attempts to formalize professional development training sometimes have generated resistance in favor of essentially an apprenticeship model with informal, ad hoc approaches to how graduate students and postdoctoral fellows learn how to become professional scientists.

One challenge facing the research enterprise is that informal, ad hoc approaches to scientific professionalism do not ensure that the core values and guiding norms of science are adequately inculcated and sustained. This has become increasingly clear as the changes in the research environment described in Chapter 3 have emerged and taken hold. Indeed, the apparent inadequacy of these older forms of training to the task of socializing and training individuals into responsible research practices is a recurring theme of this report.

Individual scientists work within a much broader system that profoundly influences the integrity of research results. This system, described briefly in Chapter 1, is characterized by a massive, interconnected web of relationships among researchers, employing institutions, public and private funders, and journals and professional societies. This web comprises unidirectional and bidirectional obligations and responsibilities between the parts of the system. The system is driven by public and private investments and results in various outcomes or products, including research results, various uses of those results, and trained students. However, the system itself has a dynamic that shapes the actions of everyone involved and produces results that reflect the functioning of the system. Because of the large number of relationships between the many players in the web of responsibility, features of one set of relationships may affect other parts of the web. These interdependencies complicate the task of devising interventions and structures that support and encourage the responsible conduct of research.

## THE CORE VALUES OF RESEARCH

The integrity of research is based on the foundational core values of science. The research system could not operate without these shared values that shape the behaviors of all who are involved with the system. Out of these values arise the web of responsibilities that make the system cohere and make scientific knowledge reliable. Many previous guides to responsible conduct in research have identified and described these values (CCA, 2010; ESF-ALLEA, 2011; IAC-IAP, 2012; ICB, 2010; IOM-NRC, 2002). This report emphasizes six values that are most influential in shaping the norms that constitute research practices and relationships and the integrity of science:

- Objectivity
- Honesty
- Openness
- Accountability
- Fairness
- Stewardship

This chapter examines each of these six values in turn to consider how they shape, and are realized in, research practices.

The first of the six values discussed in this report—objectivity—describes the attitude of impartiality with which researchers should strive to approach their work. The next four values—honesty, openness, accountability, and fairness—describe relationships among those involved in the research enterprise. The final value—stewardship—involves the relationship between members of the research enterprise, the enterprise as a whole, and the broader society within which the enterprise is situated. Although we discuss stewardship last, it is an essential value that perpetuates the other values.

## Objectivity

The hallmark of scientific thinking that differentiates it from other modes of human inquiry and expression such as literature and art is its dedication to rational and empirical inquiry. In this context, *objectivity* is central to the scientific worldview. Karl Popper (1999) viewed scientific objectivity as consisting of the freedom and responsibility of the researcher to (1) pose refutable hypotheses, (2) test the hypotheses with the relevant evidence, and (3) state the results clearly and unambiguously to any interested person. The goal is reproducibility, which is essential to advancing knowledge through experimental science. If these steps are followed diligently, Popper suggested, any reasonable second researcher should be able to follow the same steps to replicate the work.

Objectivity means that certain kinds of motivations should not influence a researcher's action, even though others will. For example, if a researcher in an experimental field believes in a particular hypothesis or explanation of a phenomenon, he or she is expected to design experiments that will test the hypothesis. The experiment should be designed in a way that allows the possibility for the hypothesis to be disconfirmed. Scientific objectivity is intended to ensure that scientists' personal beliefs and qualities—motivations, position, material interests, field of specialty, prominence, or other factors—do not introduce biases into their work.

As will be explored in later chapters, in practice it is not that simple. Human judgment and decisions are prone to a variety of cognitive biases and systematic errors in reasoning. Even the best scientific intentions are not always sufficient to ensure scientific objectivity. Scientific objectivity can be compromised acci-

dentally or without recognition by individuals. In addition, broader biases of the reigning scientific paradigm influence the theory and practice of science (Kuhn, 1962). A primary purpose of scientific replication is to minimize the extent to which experimental findings are distorted by biases and errors. Researchers have a responsibility to design experiments in ways that any other person with different motivations, interests, and knowledge could trust the results. Modern problems related to reproducibility are explored later in the report.

In addition, objectivity does not imply or require that researchers can or should be completely neutral or disinterested in pursuing their work. The research enterprise does not function properly without the organized efforts of researchers to convince their scientific audiences. Sometimes researchers are proven correct when they persist in trying to prove theories in the face of evidence that appears to contradict them.

It is important to note, in addition, Popper's suggestion that scientific objectivity consists of not only responsibility but *freedom*. The scientist must be free from pressures and influences that can bias research results. Objectivity can be compromised when institutional expectations, laboratory culture, the regulatory environment, or funding needs put pressure on the scientist to produce positive results or to produce them under time pressure. Scientists and researchers operate in social contexts, and the incentives and pressures of those contexts can have a profound effect on the exercise of scientific methodology and a researcher's commitment to scientific objectivity.

Scientific objectivity also must coexist with other human motivations that challenge it. As an example of such a challenge, a researcher might become biased in desiring definitive results evaluating the validity of high-profile theories or hypotheses that their experiments were designed to support or refute. Both personal desire to obtain a definitive answer and institutional pressures to produce "significant" conclusions can provide strong motivation to find definitive results in experimental situations. Dedication to scientific objectivity in those settings represents the best guard against scientists finding what they desire instead of what exists. Institutional support of objectivity at every level—from mentors, to research supervisors, to administrators, and to funders—is crucial in counterbalancing the very human tendency to desire definitive outcomes of research.

### Honesty

A researcher's freedom to advance knowledge is tied to his or her responsibility to be *honest*. Science as an enterprise producing reliable knowledge is based on the assumption of honesty. Science is predicated on agreed-upon systematic procedures for determining the empirical or theoretical basis of a proposition. Dishonest science violates that agreement and therefore violates a defining characteristic of science.

Honesty is the principal value that underlies all of the other relationship val-

ues. For example, without an honest foundation, realizing the values of openness, accountability, and fairness would be impossible.

Scientific institutions and stakeholders start with the assumption of honesty. Peer reviewers, granting agencies, journal editors, commercial research and development managers, policy makers, and other players in the scientific enterprise all start with an assumption of the trustworthiness of the reporting scientist and research team. Dishonesty undermines not only the results of the specific research but also the entire scientific enterprise itself, because it threatens the trustworthiness of the scientific endeavor.

Being honest is not always straightforward. It may not be easy to decide what to do with outlier data, for example, or when one suspects fraud in published research. A single outlier data point may be legitimately interpreted as a malfunctioning instrument or a contaminated sample. However, true scientific integrity requires the disclosure of the exclusion of a data point and the effect of that exclusion unless the contamination or malfunction is documented, not merely conjectured. There are accepted statistical methods and standards for dealing with outlier data, although questions are being raised about how often these are followed in certain fields (Thiese et al., 2015).

Dishonesty can take many forms. It may refer to out-and-out fabrication or falsification of data or reporting of results or plagiarism. It includes such things as misrepresentation (e.g., avoiding blame, claiming that protocol requirements have been followed when they have not, or producing significant results by altering experiments that have been previously conducted), nonreporting of phenomena, cherry-picking of data, or overenhancing pictorial representations of data. Honest work includes accurate reporting of what was done, including the methods used to do that work. Thus, dishonesty can encompass lying by omission, as in leaving out data that change the overall conclusions or systematically publishing only trials that yield positive results. The "file drawer" effect was first discussed almost 40 years ago; Robert Rosenthal (1979) presented the extreme view that "journals are filled with the 5 percent of the studies that show Type I errors, while the file drawers are filled with the 95 percent of the studies that show non-significant results." This hides the possibility of results being published from 1 significant trial in an experiment of 100 trials, as well as experiments that were conducted and then altered in order to produce the desired results. The file drawer effect is a result of publication bias and selective reporting, the probability that a study will be published depending on the significance of its results (Scargle, 2000). As the incentives for researchers to publish in top journals increase, so too do these biases and the file drawer effect.

Another example of dishonesty by omission is failing to report all funding sources where that information is relevant to assessing potential biases that might influence the integrity of the work. Conversely, dishonesty can also include reporting of nonexistent funding sources, giving the impression that the research

was conducted with more support and so may have been more thorough than in actuality.

Beyond the individual researcher, those engaged in assessing research, whether those who are funding it or participating in any level of the peer review process, also have fundamental responsibilities of honesty. Most centrally, those assessing the quality of science must be honest in their assessments and aware of and honest in reporting their own conflicts of interest or any cognitive biases that may skew their judgment in self-serving ways. There is also a need to guard against unconscious bias, sometimes by refusing to assess work even when a potential reviewer is convinced that he or she can be objective. Efforts to protect honesty should be reinforced by the organizations and systems within which those assessors function. Universities, research organizations, journals, funding agencies, and professional societies must all work to hold each other to honest interactions without favoritism and with potentially biasing factors disclosed.

## Openness

*Openness* is not the same as honesty, but it is predicated on honesty. In the scientific enterprise, openness refers to the value of being transparent and presenting all the information relevant to a decision or conclusion. This is essential so that others in the web of the research enterprise can understand why a decision or conclusion was reached. Openness also means making the data on which a result is based available to others so that they may reproduce and verify results or build on them. In some contexts, openness means listening to conflicting ideas or negative results without allowing preexisting biases or expectations to cloud one's judgment. In this respect, openness reinforces objectivity and the achievement of reliable observations and results.

Openness is an ideal toward which to strive in the research enterprise. It almost always enhances the advance of knowledge and facilitates others in meeting their responsibilities, be it journal editors, reviewers, or those who use the research to build products or as an input to policy making. Researchers have to be especially conscientious about being open, since the incentive structure within science does not always explicitly reward openness and sometimes discourages it. An investigator may desire to keep data private to monopolize the conclusions that can be drawn from those data without fear of competition. Researchers may be tempted to withhold data that do not fit with their hypotheses or conclusions. In the worst cases, investigators may fail to disclose data, code, or other information underlying their published results to prevent the detection of fabrication or falsification.

Openness is an ideal that may not always be possible to achieve within the research enterprise. In research involving classified military applications, sensitive personal information, or trade secrets, researchers may have an obligation not to disseminate data and the results derived from those data. Disclosure of results

and underlying data may be delayed to allow time for filing a patent application. These sorts of restrictions are more common in certain research settings—such as commercial enterprises and government laboratories—than they are in academic research institutions performing primarily fundamental work. In the latter, openness in research is a long-held principle shared by the community, and it is a requirement in the United States to avoid privileged access that would undermine the institution's nonprofit status and to maintain the fundamental research exclusion from national security-based restrictions.

As the nature of data changes, so do the demands of achieving openness. For example, modern science is often based on very large datasets and computational implementations that cannot be included in a written manuscript. However, publications describing such results could not exist without the data and code underlying the results. Therefore, as part of the publication process, the authors have an obligation to have the available data and commented code or pseudocode (a high-level description of a program's operating principle) necessary and sufficient to re-create the results listed in the manuscript. Again, in some situations where a code implementation is patentable, a brief delay in releasing the code in order to secure intellectual property protection may be acceptable. When the resources needed to make data and code available are insufficient, authors should openly provide them upon request. Similar considerations apply to such varied forms of data as websites, videos, and still images with associated text or voiceovers.

## Accountability

Central to the functioning of the research enterprise is the fundamental value that members of the community are responsible for and stand behind their work, statements, actions, and roles in the conduct of their work. At its core, accountability implies an obligation to explain and/or justify one's behavior. Accountability requires that individuals be willing and able to demonstrate the validity of their work or the reasons for their actions. Accountability goes hand in hand with the credit researchers receive for their contributions to science and how this credit builds their reputations as members of the research enterprise. Accountability also enables those in the web of relationships to rely on work presented by others as a foundation for additional advances.

Individual accountability builds the trustworthiness of the research enterprise as a whole. Each participant in the research system, including researchers, institutional administrators, sponsors, and scholarly publishers, has obligations to others in the web of science and in return should be able to expect consistent and honest actions by others in the system. Mutual accountability therefore builds trust, which is a consequence of the application of the values described in this report.

The purpose of scientific publishing is to advance the state of knowledge through examination by peers who can assess, test, replicate where appropriate, and build on the work being described. Investigators reporting on their work thus

must be accountable for the accuracy of their work. Through this accountability, they form a compact with the users of their work. Readers should be able to trust that the work was performed by the authors as described, with honest and accurate reporting of results. Accountability means that any deviations from the compact would be flagged and explained. Readers then could use these explanations in interpreting and evaluating the work.

Investigators are accountable to colleagues in their discipline or field of research, to the employer and institution at which the work is done, to the funders or other sponsors of the research, to the editors and institutions that disseminate their findings, and to the public, which supports research in the expectation that it will produce widespread benefits. Other participants in the research system have other forms of accountability. Journals are accountable to authors, reviewers, readers, the institutions they represent, and other journals (for the reuse of material, violation of copyright, or other issues of mutual concern). Institutions are accountable to their employees, to students, to the funders of both research and education, and to the communities in which they are located. Organizations that sponsor research are accountable to the researchers whose work they support and to their governing bodies or other sources of support, including the public. These networks of accountability support the web of relationships and responsibilities that define the research enterprise.

The accountability expected of individuals and organizations involved with research may be formally specified in policies or regulations. Accountability under institutional research misconduct policies, for example, could mean that researchers will face reprimand or other corrective actions if they fail to meet their responsibilities.

While responsibilities that are formally defined in policies or regulations are important to accountability in the research enterprise, responsibilities that may not be formally specified should also be included in the concept. For example, senior researchers who supervise others are accountable to their employers and the researchers whom they supervise to conduct themselves as professionals, as this is defined by formal organizational policies. On a less formal level, research supervisors are also accountable for being attentive to the educational and career development needs of students, postdoctoral fellows, and other junior researchers whom they oversee. The same principle holds for individuals working for research institutions, sponsoring organizations, and journals.

### Fairness

The scientific enterprise is filled with professional relationships. Many of them involve judging others' work for purposes of funding, publication, or deciding who is hired or promoted. Being fair in these contexts means making professional judgments based on appropriate and announced criteria, including processes used to determine outcomes. Fairness in adhering to explicit criteria

and processes reinforces a system in which the core values can operate and trust among the parties can be maintained.

Fairness takes on another dimension in designing criteria and evaluation mechanisms. Research has demonstrated, for example, that grant proposals in which reviewers were blinded to applicant identity and institution receive systematically different funding decisions compared with the outcomes of unblinded reviews (Ross et al., 2006). Truly blinded reviews may be difficult or impossible in a small field. Nevertheless, to the extent possible, the criteria and mechanisms involved in evaluation must be designed so as to ensure against unfair incentive structures or preexisting cultural biases. Fairness is also important in other review contexts, such as the process of peer reviewing articles and the production of book reviews for publication.

Fairness is a particularly important consideration in the list of authors for a publication and in the citations included in reports of research results. Investigators may be tempted to claim that senior or well-known authors played a larger role than they actually did so that their names may help carry the paper to publication and readership. But such a practice is unfair both to the people who actually did the work and to the honorary author, who may not want to be listed prominently or at all. Similarly, nonattribution of credit for contributions to the reported work or careless or negligent crediting of prior work violates the value of fairness. Best practices in authorship, which are based on the value of fairness, honesty, openness, and accountability, are discussed further in Chapter 9.

Upholding fairness also requires researchers to acknowledge those whose work contributed to their advances. This is usually done through citing relevant work in reporting results. Also, since research is often a highly competitive activity, sometimes there is a race to make a discovery that results in clear winners and losers. Sometimes two groups of researchers make the same discovery nearly simultaneously. Being fair in these situations involves treating research competitors with generosity and magnanimity.

The importance of fairness is also evident in issues involving the duty of care toward human and animal research subjects. Researchers often depend on the use of human and animal subjects for their research, and they have an obligation to treat those subjects fairly—with respect in the case of human subjects and humanely in the case of laboratory animals. They also have obligations to other living things and to those aspects of the environment that affect humans and other living things. These responsibilities need to be balanced and informed by an appreciation for the potential benefits of research.

### Stewardship

The research enterprise cannot continue to function unless the members of that system exhibit good stewardship both toward the other members of the system and toward the system itself. Good stewardship implies being aware of

and attending carefully to the dynamics of the relationships within the lab, at the institutional level, and at the broad level of the research enterprise itself. Although we have listed stewardship as the final value in the six we discuss in this report, it supports all the others. Here we take up stewardship within the research enterprise but pause to acknowledge the extension of this value to encompass the larger society.

One area where individual researchers exercise stewardship is by performing service for their institution, discipline, or the broader research enterprise that may not necessarily be recognized or rewarded. These service activities include reviewing, editing, serving on faculty committees, and performing various roles in scientific societies. Senior researchers may also serve as mentors to younger researchers whom they are not directly supervising or formally responsible for. At a broader level, researchers, institutions, sponsors, journals, and societies can contribute to the development and updating of policies and practices affecting research. As will be discussed in Chapter 9, professional societies perform a valuable service by developing scientific integrity policies for their fields and keeping them updated. Individual journals, journal editors, and member organizations have contributed by developing standards and guidelines in areas such as authorship, data sharing, and the responsibilities of journals when they suspect that submitted work has been fabricated or plagiarized.

Stewardship also involves decisions about support and influences on science. Some aspects of the research system are influenced or determined by outside factors. Public demand, political considerations, concerns about national security, and even the prospects for our species' survival can inform and influence decisions about the amount of public and private resources devoted to the research enterprise. Such forces also play important roles in determining the balance of resources invested in various fields of study (e.g., both among and within federal agencies), as well as the balance of effort devoted to fundamental versus applied work and the use of various funding mechanisms.

In some cases, good stewardship requires attending to situations in which the broader research enterprise may not be operating optimally. Chapter 6 discusses issues where problems have been identified and are being debated, such as workforce imbalances, the poor career prospects of academic researchers in some fields, and the incentive structures of modern research environments.

Stewardship is particularly evident in the commitment of the research enterprise to education, both of the next generation of researchers and of individuals who do not expect to become scientists. In particular, Chapter 10 discusses the need to educate all members of the research enterprise in the responsible conduct of research. Education is one way in which engaging in science provides benefits both to those within the research system and to the general public outside the system.

## A DEFINITION OF RESEARCH INTEGRITY

Making judgments about definitions and terminology as they relate to research integrity and breaches of integrity is a significant component of this committee's statement of task. Practicing integrity in research means planning, proposing, performing, reporting, and reviewing research in accordance with the values described above. These values should be upheld by research institutions, research sponsors, journals, and learned societies as well as by individual researchers and research groups. General norms and specific research practices that conform to these values have developed over time. Sometimes norms and practices need to be updated as technologies and the institutions that compose the research enterprise evolve. There are also disciplinary differences in some specific research practices, but norms and appropriate practices generally apply across science and engineering research fields. As described more fully in Chapter 9, best practices in research are those actions undertaken by individuals and organizations that are based on the core values of science and enable good research. They should be embraced, practiced, and promoted.

# 3

# Important Trends and Challenges in the Research Environment

*By working collaboratively, researchers can hope to answer questions never addressed before, including those with substantial influence on society. At the same time, today's international, interdisciplinary, team-oriented, and technology-intensive research has created an environment more fraught with the potential for error and distortion.*

—Indira Nath and Ernst-Ludwig Winnacker (2012)

***Synopsis:*** *A number of the elements in the research environment that were identified in the early 1990s as perhaps problematic for ensuring research integrity and maintaining good scientific practices have generally continued along their long-term trend lines, including the size and scope of the research enterprise, the complexity of collaboration, the growth of regulatory requirements, and the importance of industry sponsorship and entrepreneurial research. Several important new trends that were not examined in the 1992* Responsible Science *report have also emerged, including the pervasive and growing importance of information technology in research, the globalization of research, and the increasing relevance of knowledge generated in certain fields to policy issues and political debates. These changes—the growing importance of information technology in particular—have led to important shifts in the institutions that support and underlie the research enterprise, such as scholarly publishing. They also have important implications for the ability of researchers, research institutions, journals, and sponsors to foster integrity and prevent research misconduct and detrimental research practices.*

The 1992 report *Responsible Science: Ensuring the Integrity of the Research Process* devoted a chapter to describing the contemporary research environment and outlining the most important changes that had occurred over the previous decades (NAS-NAE-IOM, 1992). *Responsible Science* also described several additional features of the U.S. research scene of the early 1990s that had become the subject of discussion and concern due to possible negative impacts on the research environment, including research integrity (NAS-NAE-IOM, 1992). This chapter will first explore the research environment issues identified in 1992—except for the reward system in science, which is covered in Chapter 6—and describe trends over the past two decades. The second part of the chapter will

explore several important shifts in the research environment that have appeared since 1992 and were not considered in *Responsible Science*. These shifts carry several important implications for research integrity.

## HOW RESEARCH ENVIRONMENT ISSUES IDENTIFIED IN *RESPONSIBLE SCIENCE* HAVE EVOLVED SINCE THE EARLY 1990s

### Size and Scope of the Research Enterprise

The 1992 report's overview described growth in the size and scope of the research enterprise. The report observed that research in the pre–World War II United States—academic research in particular—was a mostly small-scale avocation of individual scientists, supported by limited funding from industry, government, and foundations. Following the significant wartime contributions of research efforts such as MIT's Radiation Laboratory, federal support for science and engineering research increased rapidly. By 1991, research and development (R&D) was a $160 billion (current dollars) enterprise in the United States, employing about 744,000 people in industrial, academic, and governmental laboratories and producing more than 140,000 research articles annually (NSB, 1996, 2014b; OECD, 2015).

Over the following two decades, the enterprise has continued to grow, with U.S. R&D totaling $456 billion in 2013, R&D employment rising to about 1,252,000, and the number of published research articles reaching more than 412,000 (NSB, 2014b, 2016; OECD, 2015). The 1992 report paid particular attention to the growth in academic research and federal support, and this growth has continued. Between 1991 and 2014, academic R&D grew from around $17.5 billion to $67.1 billion, with federal support constituting 60–75 percent of the total (NSB, 2016).[1] The number of science, engineering, and health doctorate holders employed in academia rose from 211,000 in 1991 to almost 309,000 in 2013 (NSB, 2016). The number of PhDs awarded in science and engineering more than doubled, from approximately 19,000 in 1988 to almost 37,000 in 2013, with an increasing percentage of these doctorate recipients going to work outside academia (NSB, 2016).

The 1992 *Responsible Science* report raised the concern that the increased size of the research enterprise might put stresses on key capabilities, such as the "overall workload associated with critical evaluation" (NAS-NAE-IOM, 1992). The number and capacity of effective peer reviewers might not be keeping pace with the relentless growth in manuscripts and proposals. Concerns also have been raised about the increasing use of bibliometric-based metrics in evaluating

[1] From 2010, the total includes academic R&D outside of science and engineering, which adds several billion dollars.

research as a substitute or replacement for expert judgment (P. B. Lowry et al., 2012).

## Complexity of Collaboration

*Responsible Science* described the growth of collaborative research after World War II, which has continued since the early 1990s. In contrast to earlier times, when articles with more than four co-authors and work involving more than one laboratory or research institution were rare, collaborative research of various types is now very common. The number of authors listed on articles is only one measure of collaboration, but it clearly reveals the overall trend. In an analysis of approximately 20 million research articles published since 1955 and 2 million patents registered since 1975, the number of authors on scientific papers grew from an average of 1.9 in 1955 to 3.5 in 2000 (Wuchty et al., 2007). At the same time, single-author articles are becoming less common, constituting only about 11 percent of the total in 2012 (King, 2013).

Several factors are driving the trend toward larger-scale research in general and in specific fields (Stephan, 2012a). These include the need for more elaborate and expensive equipment and the often related requirement for a variety of specialized skills and knowledge. These characteristics of "big science" have long been a given in fields such as high-energy physics and astronomy, in the form of particle accelerators such as the Large Hadron Collider and modern telescopes. They have become more prominent recently in many areas of the life sciences as well. In describing the results of large life sciences research projects such as the Human Genome Project and ENCODE (Encyclopedia of DNA Elements), former *Science* editor-in-chief Bruce Alberts (2012) noted that "the increased efficiency of data production by such projects is impressive." In addition, as will be discussed in more detail below, the information technology revolution has radically lowered the costs of communication and collaboration of all types, including research collaboration.

Another factor contributing to the growth of team research has been an increase in the amount of interdisciplinary research. Interdisciplinary research efforts have continued to grow in importance and are extremely diverse (Derrick et al., 2012). Interdisciplinary teams can range from local and informal to transnational and highly structured. They can be composed largely or entirely of researchers accustomed to working within a disciplinary framework, or they can consist partly or wholly of researchers who have been educated and have worked in interdisciplinary fields. Integration of knowledge from multiple disciplines can occur within the mind of a single person or through the collaborative efforts of a large team. For example, with the advent of "big data" and computational science, statisticians are increasingly included on projects where researchers have collected domain-specific data that they do not have the expertise to analyze. Interdisciplinary research is often focused on problems that have important so-

cietal implications. One current example of a growing interdisciplinary field is synthetic biology, which seeks a fundamental understanding of the workings of living systems along with the capability of re-creating living systems for a variety of applications in areas such as medicine and the environment. Synthetic biology research involves "biologists of many specialties, engineers, physicists, computer scientists, and others" (NRC, 2010).

According to one analysis of trends in interdisciplinary research in six research fields, the growth of interdisciplinarity has been modest—about 5 percent—even as the number of authors per article has grown by 75 percent (Porter and Rafols, 2009). This study found that the number of disciplines cited by papers in these six fields—mathematics, neurosciences, electrical and electronic engineering, biotechnology and applied microbiology, research and experimental medicine, and atomic, molecular, and chemical physics—has increased, but the distribution of citations is within neighboring research areas and has only slightly broadened. According to the authors, "These findings suggest that science is indeed becoming more interdisciplinary, but in small steps—drawing mainly from neighboring fields and only modestly increasing the connections to distant cognitive areas."

Collaborative science requires that researchers focus at least some attention on coordination and interaction, which in theory might detract from the time and effort devoted to research. Yet Wuchty et al. (2007) found that multiauthor teams produced more highly cited work in each broad area of research and at each point in time. In addition, though solo authors in 1955 were more likely to produce papers that were highly cited, suggesting that these papers reported on the most influential concepts, results, or technologies, teams are more likely to produce highly cited papers today. As the authors wrote, "solo authors did produce the papers of singular distinction in science and engineering and social science in the 1950s, but the mantle of extraordinarily cited work has passed to teams by 2000."

As more researchers work collaboratively and as the size of teams grows, the relationships among team members can become more complex. Team members can be at different research institutions and have different disciplinary backgrounds. Teams can contain researchers at all stages of their careers, from undergraduate and graduate students involved in research to senior researchers. The diversity and geographic spread of people involved in teams can create opportunities for miscommunication, misunderstandings, unrealistic expectations, and unresolved disputes. Whether these opportunities account for part of the increase in reports of undesirable research practices is unclear, but they can make the research environment more complicated and difficult than when teams were smaller, colocated more regularly, and more homogeneous in terms of discipline or nationality.

As research projects are undertaken by larger groups that bring together a greater diversity of expertise, encompass a broader range of disciplines, and strive for a greater degree of synthesis, the potential for misunderstandings can grow. Coordination of research inevitably becomes more complex, and the members

of a team may have less familiarity with the discipline-specific practices of other team members, making it more difficult for each collaborator to check and verify the work done by others. As the number of collaborators increases, there is more scope for disagreements over the allocation of credit. It becomes much more challenging to reward and recognize individual contributions, which has a big impact on junior researchers in particular. In addition, the mentoring of students in responsible research practices can become more impersonal and generic. The mental model of graduate education and training in which mentors work closely with graduate students and are able to take the time and effort to ensure that mentees understand the rules and can follow them may describe a smaller and smaller part of the research enterprise. Interdisciplinary work increases the possibility that the standards and expectations of different fields may come into conflict.

### Regulation and Accountability

The 1992 report also noted that research activities were "increasingly subject to government regulations and guidelines that impose financial and administrative requirements" in areas such as laboratory safety, human subjects protection, drug-free workplace assurance, laboratory animal care, and the research use of recombinant DNA and toxic and radioactive materials. Along with the relatively new requirements and regulations related to research misconduct, the development of which is covered in Chapter 4 of this report, ensuring compliance with these expanding regulatory requirements had resulted in an expansion of administrative and oversight functions and staff at universities and required increasing time and attention from investigators. As an increasing percentage of faculty time goes toward fulfilling the requirements of various regulations and reporting requirements, research-related tasks such as mentoring and checking the work of subordinates may be shortchanged.

The administrative and regulatory compliance burden on research institutions and researchers remains significant. For example, respondents to a 2012 survey of 13,453 principal investigators undertaken by the Federal Demonstration Partnership estimated that, on average, 42 percent of the time they spent working on federally funded research projects was devoted to meeting regulatory and administrative requirements (Schneider et al., 2012). According to the survey results, areas of regulation where compliance is particularly time consuming include those related to finances, personnel, and effort reporting. In 2014 the National Science Board issued a report that analyzes the regulatory compliance burden on faculty and makes recommendations for how it might be reduced (NSB, 2014c). A 2016 National Academies report evaluated current approaches to regulating academic research and made recommendations for achieving the goals of regulation while reducing financial and time burdens on institutions and faculty (National Academies of Sciences, Engineering, and Medicine, 2016).

## Industry-Sponsored Research and Other Research Aimed at Commercialization

Increasingly, the scientific enterprise has been recognized not only as a place to expand knowledge but also as an engine for the creation of new products, novel therapies for disease, improved technologies, and new industries and jobs. To quote President Obama (2009b), "scientific innovation offers us a chance to achieve prosperity." The economic potential of science, however, also offers unique challenges to the responsible conduct of research, which were described in *Responsible Science*. These challenges can be seen in scientific research conducted in an industrial setting, scientific research conducted in university and research institutions in collaboration with industry, and university research that leads to entrepreneurial efforts by the researchers, requiring that they integrate both within themselves and in their professional behavior often divergent cultural understandings about the nature, purposes, and outcomes of research. These challenges include the potential of economic incentives to introduce scientific bias, the perception of conflict of interest due to economic incentives, and the potential effect of intellectual property protection on the timely dissemination of knowledge.

Industry funds and conducts a substantial amount of research in the United States. For both basic and applied research, as defined by the National Science Foundation, industry conducts 40 percent of the U.S. total (NSB, 2016). Even considering just basic research, industry conducts approximately 24 percent, almost 90 percent of which it funds itself. Unlike academic research, corporate research is often driven by the needs of a company to remain financially solvent and to be accountable to shareholders. Corporate researchers often exist under hierarchical chains of supervision where management maintains greater control over the research process.

Only a fraction of the results of industry-funded research is published in the scientific and engineering literature and is thereby submitted to formal peer review. Of the articles published in 2013, authors from industry accounted for only 6 percent of the total, and that percentage has been declining over the past two decades (NSB, 2014). This can be a product of the need to preserve intellectual property interests for trade secrets and obtaining patents. One consequence is that the knowledge gained in such research may not be widely disseminated or evaluated through the peer review process. This is not to say that such industry research is not of high quality or is not carefully reviewed. Companies can have strict protocols regarding the collection, documentation, and storage of data, particularly when there are strong regulatory or economic reasons to do so. Checking mechanisms may be built into industrial research to verify especially critical results (Williams, 2012). And, as with all research, the use of research results in subsequent activities—including the production of commercial products—provides further checks on the validity of results.

However, both industrial research and industry-sponsored research in aca-

demic settings have been found to occasionally show signs of both unintentional and intentional bias.[2] For example, one might observe bias in the lack of publication of results with negative consequences for the profitability of a product or in the restriction of published findings to those that reflect positively on a product. An extreme case is the tobacco industry, which undertook a systematic effort over the course of decades to obscure the harmful effects of smoking (Proctor, 2011). Other examples include episodes of alleged ghostwriting in some medical research, including the Paxil case described in Appendix D and also discussed in Chapter 7. Such research tarnishes all other research by demonstrating that research agendas and techniques can be manipulated so severely as to subvert truth to other interests. Many journals have moved to reporting the financial interests of authors, whether the work has an industry sponsor or not, so that readers are made aware of the potential for bias.

In addition to collaborations with established industries, academic institutions have increasingly encouraged entrepreneurship and innovation for commercialization, particularly since the passage of the Bayh-Dole Act in 1980, which allowed institutions to hold patents on innovations produced with federal funding. Having seen the success of academic research products such as Gatorade and the Google search algorithm patent in generating revenue, institutions may hope that their researchers can achieve similar results. For fiscal 2011 the Association of University Technology Managers reported that the 186 institutions responding to its annual survey earned a total of $1.5 billion in running royalty income, executed 4,899 licenses, created 591 commercial products, and formed 671 start-up companies from their research (AUTM, 2012).

One result of the commercialization of university-generated technology is that the need to manage possible conflicts of interest has become an important issue in academic settings. A 2009 Institute of Medicine report explores the issue of institutional conflict of interest in more detail (IOM, 2009). Individual conflicts of interest exist if the investigator is also the founder of a company conducting research or has a significant monetary stake in the research. This can also apply to an institution if it owns part of a company or has a financial stake in a faculty member's research findings. Under the U.S. Financial Conflict of Interest (FCOI) policy, research funded by the Public Health Service requires institutions to maintain and enforce a FCOI policy; manage, reduce, or eliminate identified conflicts; report identified conflicts, the value of the conflicts, and a management plan to the Public Health Service Awarding Component; and publish significant financial interests of any personnel involved in the research on a publicly accessible website (HHS, 2011b). Currently, the Department of Health and Human Services does not have institutional regulations in the same manner as investigator FCOI regulations (required disclosure of FCOIs). Strengthened institutional FCOI regulations have been considered, but there is a need for further and separate consideration.

[2] This is not meant to imply that research that is not sponsored by industry is necessarily unbiased.

The National Science Foundation policy is consistent with that of the Department of Health and Human Services. Regulations of individual financial conflicts of interest are further discussed in Chapter 7 and are also addressed in the context of best practices in Chapter 9.

Additional individual conflicts of interest, or secondary interests, can also affect a research study, including political biases, white hat bias, commitment conflicts, career considerations, and favors to others (IOM, 2009; Lesser et al., 2007). A political opinion, bias, or long-standing scientific viewpoint toward one position or another may influence the interpretation of findings, despite contradictory evidence (Lesser et al., 2007). Similarly, white hat bias, or "bias leading to distortion of information in the service of what may be perceived to be righteous end," also has the potential to influence conclusions (Cope and Allison, 2010). An example of a conflict of commitment would be a principal investigator who does not have the time to perform all the duties for which he or she is responsible, such as securing funding, setting the overall direction for research in a lab, administrative responsibilities, and adequately supervising graduate students and postdocs. Secondary interests are rarely regulated, as they are considered a lesser incentive than financial interests.

Closer linkages between research and commercialization have introduced the possibility of financial gain from research more widely across the enterprise. This can pose challenges in terms of defining appropriate behavior and establishing guidelines for dealing with conflicts of interest, and it can complicate collaborations among individual researchers and among organizations.

## MAJOR CHANGES IN THE RESEARCH ENVIRONMENT SINCE 1992

### Information Technologies in Research

The continued exponential rise in the power of information and computing technologies has had a dramatic impact on research across many disciplines. These technologies have not only increased the speed and scope of research but have made it possible to conduct investigations that were not possible before. Information technology advances have enabled new forms of inquiry such as those based on numerical simulation of physical and biological systems and the analysis of massive datasets to detect and assess the nature of relationships that otherwise would go unseen.

The contrast in computing capabilities since the publication of *Responsible Science* is especially stark. In 1992, use of e-mail was less than a decade old, and the World Wide Web had just been invented and was not widely known. Three-and-a-half-inch floppy disks for data storage had replaced 5-1/4-inch disks just a few years before. People made telephone calls on landlines, used letters to communicate in writing, and circulated preprints via the postal system. For

young researchers, the circumstances in which research was conducted in 1992 are almost entirely foreign.

One effect of information technologies in many areas of research has been to introduce intermediate analyses of considerable complexity between the "raw" data gathered by sensors and observations, and produced by data-creating devices such as DNA sequencers, and the results of research. Re-creating the steps from data to results can be impossible without a detailed knowledge of data production and analyzing software, which sometimes is dependent on the particular computer on which the software runs. This intermediate analysis complicates the replication of scientific results and can create opportunities to manipulate analyses so as to achieve desired results, as well as undermine the ability of others to validate findings.

Digital technologies can pose other temptations for researchers to violate the standards of scientific practice. For example, the manipulation of images using image-processing software has caused many journals to implement spot checks and other procedures to guard against falsification. The inappropriate application of statistical packages can lead to greater confidence in the results than is warranted. Data-mining techniques can generate false positives and spurious correlations. In many fields, the development of standards governing the application of technology in the derivation of research results remains incomplete even as continuing technological advances raise new issues. In a recent paper, two prominent biologists wrote, "Although scientists have always comforted themselves with the thought that science is self-correcting, the immediacy and rapidity with which knowledge disseminates today means that incorrect information can have a profound impact before any corrective process can take place" (Casadevall and Fang, 2012).

The widespread utilization of information technologies in research may also introduce new sources of unintentional error and irreproducibility of results. A survey of researchers who utilize species distribution modeling software found that only 8 percent had validated the software they had chosen against other methods, with higher percentages relying on recommendations from colleagues or the reputation of the developer (Joppa et al., 2013). The latter approaches pose risks of incorrect implementation and error for the research being pursued, particularly if software is not shared or subjected to critical review. Issues surrounding irreproducibility and information technologies are discussed further in Chapter 5.

Besides affecting the conduct of research, information and communication technologies have transformed the communication of scientific results and interactions among researchers. In theory, if not always in practice, all the data contributing to a research result can now be stored electronically and communicated to interested researchers. This capability has contributed to a growing movement for much more open forms of research in which researchers work collectively on problems, often through electronic media (Nielsen, 2012). However, this trend toward greater transparency has created tasks and responsibilities for research-

ers and the research enterprise that did not previously exist, such as creating, documenting, storing, and sharing scientific software and immense databases and providing guidance in the use of these new digital objects. For example, software produced by scientists in the course of analyzing the data is often carried out as a collaborative online process. This digitization makes it easier than ever to perform very complex analyses that not only lead to new discoveries but create new problems of opacity for the peer review process. And while technology is making many aspects of research more efficient, it may also create new tasks and responsibilities that are burdensome for researchers and that they may find difficult or impossible to fulfill.

The movement toward open science has encouraged the efforts of citizen scientists who are eager to monitor, contribute to, and in some cases criticize scientific advances (Stodden, 2010). Review of scientific results from outside a research discipline can provide another check on the accuracy of results, but it also can introduce questions about the validity of findings that are not adequately grounded in knowledge of the research. Moreover, it can alter the relationship between researchers and the public in ways that require new levels of effort and sophistication among researchers involved in public outreach.

Advances in information technology are transforming the research enterprise, discipline by discipline, by changing the sorts of questions that can be addressed and the methods used to address them. There may be more opportunities to fabricate, falsify, or plagiarize, but there are also more tools to uncover such behavior. Issues related to research reproducibility and related practices are covered in Chapter 5.

## The Globalization of Research

Because knowledge passes freely across national borders, scientific research has always been an international endeavor. But this internationalization has intensified over the past two decades. Nations have realized that they cannot expect to benefit from the global research enterprise without national research systems that can absorb and build on that knowledge. As a result, they have incorporated science and technology into national plans and have established goals for increased R&D investments. They also have encouraged their own students and researchers to travel to other countries to study and work and have welcomed researchers from other countries. At the same time, private-sector companies have increased their R&D investments in other countries to take advantage of local talent, gain access to local markets, and in some cases lower their costs for labor and facilities. These and other trends, including cheaper transportation, better communications, and the spread of English as the worldwide language of science, are producing a new golden age of global science.

Once again, the trend is apparent in the author lists of scientific and engineering articles. Between 1988 and 2013, the percentage of science and engineer-

ing articles published worldwide with coauthors from more than one country increased from 8 percent to 19 percent (NSB, 2016). Also, some countries have dramatically increased their representation in the science and engineering literature. Between 1999 and 2013, the average number of science and engineering articles published by Chinese authors rose 18.9 percent annually, so that by 2013 China, with 18 percent of the total, was the world's second-largest national producer of science and engineering articles. Authors from China also increased their share of internationally coauthored articles from 5 percent to 13 percent between 2000 and 2010. Other countries that dramatically expanded their number of articles published included South Korea, India, Taiwan, Brazil, Turkey, Iran, Greece, Singapore, Portugal, Ireland, Thailand, Malaysia, Pakistan, and Tunisia, though some of these countries started from very low bases.

Another measure of the increasing internationalization of research is the number of foreign-born researchers studying and working in the United States. More than 193,000 foreign students were enrolled in U.S. graduate programs in science and engineering in 2013, and foreign-born U.S. science and engineering doctorate holders held 48 percent of postdoctoral positions in 2013 (NSB, 2016). Science and engineering doctorate holders employed in U.S. colleges and universities who were born outside the United States increased from 12 percent in 1973 to nearly 27 percent in 2013. The United States remains the destination for the largest number of foreign students at the graduate and undergraduate levels, though its share of foreign students worldwide declined from 25 percent in 2000 to 19 percent in 2013.

Internationalization offers many benefits to the research enterprise. It can speed the advance of knowledge and permit projects that could not be done by any one country working alone. It increases cooperation across borders and can contribute to a reduction in tensions between nations. It enhances the use of resources by reducing duplication of effort and by combining disparate skills and viewpoints. The experiences students and researchers gain by working in other countries are irreplaceable.

But globalization also can complicate efforts to ensure that researchers adhere to responsible research practices (Heitman and Petty, 2010). Education in the responsible conduct of research, while far from universal among U.S. science and engineering students, is nevertheless more extensive in the United States than in many other countries (Heitman et al., 2007). Codes of responsible conduct differ from country to country, despite efforts to forge greater international consensus on basic principles (ESF-ALLEA, 2011; IAC-IAP, 2012). In some countries with rapidly developing research systems, research misconduct and detrimental research practices appear to be more common than in countries with more established research systems (Altman and Broad, 2005). Students from different countries may have quite different expectations regarding such issues as conflicts of interest, the deference to be accorded instructors and mentors, the treatment of research subjects, the handling of data, and the standards for authorship. For

example, one issue often noticed with foreign students in the United States is the different standards they apply to the use of ideas and phrases from others, which can lead to problems with plagiarism (Heitman and Litewka, 2011).

As the sizes of individual national research enterprises grow and become more competitive, institutions and sponsors can experience more problems with research misconduct. Differences in national policy frameworks may constitute barriers to cross-border collaboration, but efforts are being made to harmonize or at least make these frameworks interoperable. Collaboration among researchers from different countries and cultures may expose differences in training, expectations, and values that affect behavior.

### Relevance of Research Results to Policy and Political Debates

The rapid expansion of government support for scientific research in the decades after World War II was spurred by recognition of the importance of new knowledge in meeting human needs and solving problems. Over the past few decades, the link between scientific knowledge and issues in the broader society has become ever more apparent. Science is a critical factor in public discussions of and policy decisions concerning stem cells, food safety, climate change, nuclear proliferation, education, energy production, environmental influences on health, national competitiveness, and many other issues. Although all these topics cannot be covered here, this section will describe several of the key issues affecting science, policy, and the public and how they affect (and are affected by) research integrity.

To begin with, the federal government itself performs a significant amount of research through government laboratories, some of which is published. Federal agencies that perform research generally have policies and procedures in place to investigate allegations of research misconduct in their intramural programs (see NIH, 2012a, for an example of such policies and procedures, and see Chapter 7 for a more detailed discussion).

In addition, the Obama administration led an initiative on scientific integrity in the federal government starting in 2010 (Holdren, 2010). Executive departments and agencies were instructed by the Office of Science and Technology Policy (OSTP) to develop policies that address a range of issues, including promoting a culture of scientific integrity, ensuring the credibility of government research, fostering open communication, and preventing bias from affecting how science is used in decision making or in communications with the public. The exercise is largely complete, as agencies have developed and implemented policies in response to the Office of Science and Technology Policy guidance (Grifo, 2013; OSTP, 2013).

Research also comes into play in debates and decisions over numerous contentious policy issues. Science is not the only factor in these discussions. Many considerations outside of science influence policy choices, such as personal and

political beliefs, lessons from experience, trial-and-error learning, and reasoning by analogy (NRC, 2012b). To contribute to public policy decisions, researchers must be able to separate their expertise as scientists from their views as advocates for particular public policy positions. Furthermore, they often contribute to these discussions outside the peer-reviewed literature, whether in public forums, blogs, or opinion articles in newspapers. According to the document *Responsible Conduct in the Global Research Enterprise: A Policy Report* (IAC-IAP, 2012), "Researchers should resist speaking or writing with the authority of science or scholarship on complex, unresolved topics outside their areas of expertise. Researchers can risk their credibility by becoming advocates for public policy issues that can be resolved only with inputs from outside the research community."

One example of an area where science, public debate, and policy making have been closely tied and contentious in recent years is climate science. This has raised challenges for researchers and the institutions through which scientists provide policy advice. According to a recent National Research Council report, "Climate change is occurring, is very likely caused by human activities, and poses significant risks for a broad range of human and natural systems. The environmental, economic, and humanitarian risks of climate change indicate a pressing need for substantial action to limit the magnitude of climate change and to prepare to adapt to its impacts" (NRC, 2011). The global climate is a highly complex system, and there is considerable uncertainty about the timing and magnitude of climate change, the effect of measures to reduce greenhouse gas emissions from human activities, regional impacts, and many other issues. Effectively limiting greenhouse gas emissions presents economic and technological challenges and affects countries and industries differently, making policy changes by individual countries difficult. The development of the United Nations Framework Convention on Climate Change and its evolution over time illustrate the barriers to collective action on a global level.[3]

In this environment of significant uncertainty on key scientific questions, difficult policy choices, the possibility of large impacts on powerful economic interests, and highly mobilized advocacy operations on all sides of the climate change issue, the climate science community has faced challenges in maintaining its credibility and public trust as it contributes its expertise. This experience might provide lessons on what researchers and scientific institutions need to do and what they need to avoid as highly charged issues arise with important scientific components. For example, the Intergovernmental Panel on Climate Change (IPCC), which was awarded the Nobel Peace Prize in 2007, is an international body that undertakes periodic scientific assessments of climate science and constitutes the primary mechanism for scientists to inform policy makers at the global level. In November 2009 the unauthorized leak of e-mail conversations among climate researchers, a number of whom were heavily involved with the IPCC process,

[3] See http://unfcc.int./meetings/warsaw_nov2013/meeting/7649.php.

appeared to reveal a number of questionable actions, including efforts to limit or deny access to data, failure to preserve raw data, and efforts to influence the peer review practices of journals. While subsequent investigations cleared the researchers of misconduct, the "Climategate" scandal and subsequent discovery of errors in IPCC's most recent assessment raised questions about the quality and impartiality of the organization's work. A 2010 study by the InterAcademy Council recommended a number of reforms in IPCC governance and management, review processes, methods for communicating uncertainty, and transparency (IAC, 2010). One possible lesson from the recent climate change experience is that researchers, institutions, and fields whose work becomes relevant to controversial policy debates will need to consciously examine and upgrade their practices in areas such as data access and transparency (NAS-NAE-IOM, 2009a).

Recent high-profile international cases in which scientists have been criticized and even prosecuted based on their advisory activities include the statements of scientists in the aftermath of the Fukushima earthquake and tsunami in 2011, and the manslaughter convictions of seismologists whose statements were misconstrued by a government official, Bernardo De Bernardinis, to mean that there was no risk of danger immediately prior to an earthquake in L'Aquila, Italy, that killed more than 300 people (Cartlidge, 2012; Jordan, 2013; Normile, 2012). An appeals court overturned the convictions 2 years later for the six seismologists involved, but not for De Bernardinis (Cartlidge, 2014).

Other issues involving science and policy that raise questions about integrity seemingly appear in the media on a weekly basis. During 2012, controversy erupted over a University of Texas sociologist's research findings that adult children of parents who had same-sex relationships fared worse than those raised by parents who had not had same-sex relationships; his research methodologies have been severely criticized, but an institutional inquiry cleared him of research misconduct (Peterson, 2012). A federal appeals court upheld a South Dakota statute requiring doctors to tell women seeking abortions that they face an increased risk of suicide; despite extremely weak research evidence to support the statute, the court decided not to strike it down as an undue burden on abortion rights or on First Amendment grounds (Planned Parenthood Minnesota, N.D., S.D. v. Rounds, 2012). A French paper found that rats consuming genetically modified corn developed more tumors and died earlier than a control group, although food safety agencies have stated that the sample sizes were too small to reach a conclusion (Butler, 2012). And a criminal investigation of a Texas state agency established to fund research on cancer prevention and treatment revealed that some awards were made without scientific review, which led to a wave of resignations among staff and oversight board members (Berger and Ackerman, 2012). Needless to say, these cases underscore the salient role of scientific research in policy discussions.

For researchers, exercising responsibility in relations with society encompasses an increasing array of issues. For example, health and social science research in some communities, such as Native American tribes, requires adher-

ence to community rules for gaining approval. Research on people's behavior on social networking websites raises questions about how human subject protections apply. Some emerging areas of research, such as crisis mapping and monitoring, raise human rights issues (AAAS, 2012). Finally, researchers in the life sciences are being asked to exercise responsibility in the area of preventing the misuse of research and technology (IAP, 2005).

Research findings are increasingly relevant to a broader range of policy-relevant questions, raising the magnitude of possible negative consequences of research misconduct and detrimental research practices. Researchers in a variety of fields are faced with more complicated choices with ethical dimensions. In this environment, maintaining rigorous peer review processes in scientific journals is a critical task. Decisions based on science suffer when non-peer-reviewed science, or science that was not well reviewed, is used.

## TRENDS IN RESEARCH AND IMPLICATIONS FOR AUTHORSHIP

Decisions about the authorship of research publications are an important aspect of the responsible conduct of research. Although many individuals other than those who conceive of and implement a research project typically contribute to the production of successful research, authors are considered to be the person or persons who made a significant and substantial contribution to the production and presentation of the new knowledge being published. A number of the conventions and practices that constitute scientific authorship have been influenced by the trends discussed previously in this chapter. Tracing how trends in research such as globalization and technology are affecting authorship provides a useful window into how research is changing more broadly.

Authorship practices have evolved to support the development and distribution of new knowledge, engaging the powerful human motivation to discover and receive credit for discovery. Researchers are often evaluated, rightly or wrongly, by the quantity and quality of their work, as measured by the number of their publications, the prestige of the journals in which their publications appear, and how widely cited their publications are. Authorship also serves to establish accountability for published work. For example, authors are responsible for the veracity and reliability of the reported results, for ensuring that the research was performed according to relevant laws and regulations, for interacting with journal editors and staff during publication, and for defending the work following publication (Smith and Williams-Jones, 2012).

Authorship practices vary between disciplines. Professional and journal standards and policies on authorship also vary. For example, in some disciplines the names of authors are listed alphabetically, while in other disciplines names are listed in descending order of contribution. In some disciplines, senior researchers are listed last and in others they are listed first.

At least three significant factors have changed authorship practices in recent

decades. First, the degree to which researchers make use of technology and the ways in which they use technology have changed dramatically. Researchers now frequently rely on computer software and hardware for many of the processes and analyses they undertake. They rely more on sophisticated software and computer models both in the analysis and in the presentation of results. The extent to which researchers understand how these tools affect data and results is a topic of concern in 21st-century research. Second, as a result of new information and communication technologies, especially the Internet, researchers engage in much more collaboration at a distance. This facilitates national and global collaboration and can lead to larger, more broadly scoped projects. Data gathering and analysis can be parsed out to different locations, with information potentially easily accessed and shared regardless of location. Researchers are able to electronically maintain frequent contact, have group meetings, and coauthor documents. Third, as a result of software and hardware developments, huge databases of information can be gathered and used, and researchers have access to and must deal with much more information than ever before. Consequently, researchers have to manage data in new ways and may be held to higher standards of knowing and understanding other research that has been done in their area.

These changes raise a variety of challenges to researchers and the research enterprise. For example, in part because of the increased scale of research, the number of authors listed on papers in some disciplines has grown considerably. Extreme examples include the 1993 Global Utilization of Streptokinase and Tissue Plasminogen, or GUSTO, paper in the *New England Journal of Medicine*, which involved 976 authors (GUSTO Investigators, 1993), and a 1997 *Nature* article on genome sequencing that had 151 authors (Kunst et al., 1997, from Smith and Williams-Jones, 2012). The recent joint paper from the two teams collaborating on the mass estimate of the Higgs boson particle lists more than 5,000 authors (Castelvecchi, 2015). The original papers reporting the discovery of the Higgs boson had approximately 3,000 authors each (Hornyak, 2012). How can the primary author or authors be responsible for the work of a hundred individual researchers who are geographically dispersed and come from a wide range of disciplines? When an error is found or an accusation of wrongdoing is made, the problem has to be traced back to the component of the research that is called into question. In the process of tracing back the possible wrongdoing, the primary author or authors, while accountable, may not understand the area or have had much control over the researchers involved. The primary author or authors may be accountable but not blameworthy. These challenges are complicated by disciplinary differences in authorship conventions.

Chapter 7 explores the challenges to research integrity arising in the area of authorship, and Chapter 8 considers alternatives for addressing them.

# Part Two

# Research Misconduct and Detrimental Research Practices

# 4

# Context and Definitions

*In the end, a commitment to the ethical standard of truthfulness, through an understanding of its meaning to science, is essential to enhance objectivity and diminish bias. Unfortunately, the ethos of concern for scientific misconduct continues to dominate the research-ethics movement. This focus is damaging because it turns the attention to seeking and finding wrong-doers and determining punishment rather than discussing generic issues of doing the right thing, preventing harms, seeking benefits, and understanding the right-making and wrong-making characteristics of actions. The focus on scientific misconduct makes ethical issues appear synonymous with legal issues and the search for ethical understanding synonymous with carrying out an investigation.*

—S. J. Reiser (1993)

***Synopsis:*** *Integrity is essential to the functioning of the research enterprise and personally important to the vast majority of those who dedicate their lives to science. Yet research misconduct and detrimental research practices are facts of life. They must be understood and addressed. This chapter begins with a brief historical overview of misconduct in science, followed by a discussion of definitions and categories that the committee recommends for use by the research enterprise going forward. This framework retains many key aspects of the 1992 committee's work but suggests several changes.*

## HISTORICAL CONTEXT

Prominent cases of research misconduct have been uncovered regularly over the time that science has existed as an organized activity. The Piltdown Man hoax of the early 20th century is perhaps the most famous of numerous archaeological hoaxes and frauds that have continued up to recent times. In 2000, amateur archaeologist Shinichi Fujimura was found to have "discovered" artifacts that he had placed in older strata than where they had actually been found. Other fields, such as evolutionary biology, are also represented. Fraudulent work in the first half of the 20th century by Paul Kammerer and Trofim Lysenko purported to prove environmentally acquired inheritance. Questions have even been raised about the integrity of work by revered scientists from the past (Broad and Wade, 1983; Goodstein, 2010).

According to the report *Responsible Science* (NAS-NAE-IOM, 1992), "until [recently] scientists, research institutions, and government agencies relied solely on a system of self-regulation based on shared ethical principles and generally accepted research practices to ensure integrity in the research process." As discussed in Chapter 2, science and research have not had defined mechanisms for certification, licensure, and imposing penalties for unethical behavior that have developed in professions such as medicine, law, and some areas of professional engineering. Behaviors such as fabrication of research results and plagiarism might be punished by employers but were generally not subject to legal action, at least in the United States.[1]

Unethical behavior in research first emerged as a policy issue in connection with the treatment of human research subjects and laboratory animals. While ethical concerns about human subjects were first raised earlier, it was the Nazi and Japanese military experiments on prisoners during World War II that led to the development of formal international codes. The Tuskegee syphilis study by the U.S. Public Health Service (PHS) that was launched in the 1930s, but only became subject to publicity and critical examination in 1972, provided impetus for policy changes. Policies to protect human subjects and laboratory animals were adopted in the United States during the 1960s and 1970s.

A series of cases in which researchers fabricated data or plagiarized the work of others garnered considerable publicity and prompted congressional hearings in 1981 (Medawar, 1996; Rennie and Gunsalus, 2001; Steneck, 1994). Conflict-of-interest questions also began arising in this period, related to the effects of researchers benefiting from studies by being awarded stock and other rewards. Due in part to the growth of the research enterprise and the steady increase in federal funding for research, these high-profile cases of fabrication or plagiarism in publicly funded studies were seen as examples of defrauding taxpayers and resulted in congressional attention. Federal agencies began to develop policies on research misconduct during the 1980s. During the late 1980s and early 1990s, cases of alleged immunology data falsification and fabrication against pathologist Thereza Imanishi-Kari of Tufts University (a collaborator of Nobel Prize winner David Baltimore) and data falsification allegations against Mikulas Popovic and Robert Gallo at the National Institutes of Health attracted significant attention from Congress and the news media (Gold, 1993; Kaiser, 1997; Kevles, 1998). After lengthy, complicated, and controversial investigations and adjudication processes, none of the accused in these cases was found to have committed research misconduct. However, these cases provided an important impetus for federal agencies—the Department of Health and Human Services and the National Science Foundation (NSF) in particular—to regularize how allegations of research misconduct would be investigated and adjudicated by specifying the responsi-

[1] The contexts where data fabrication is subject to criminal prosecution in the United States are discussed in Chapter 7.

bilities of research institutions, the practices that constitute misconduct and are subject to corrective action, and the oversight roles of the agencies themselves.

These cases had a significant impact on the development of federal and institutional approaches to addressing misconduct. The evolution of these approaches is summarized in Table 4-1. Current approaches to addressing research misconduct and detrimental research practices are described in detail in Chapter 7.

## WHY IS A FRAMEWORK OF CONCEPTS AND DEFINITIONS OF KEY TERMS NEEDED?

Chapter 2 explored the values underlying research and the behaviors that express those values. As behaviors that violate those values, such as data fabrication, emerged as serious problems, researchers and policy makers sought to develop a framework of concepts and definitions to use in preventing, investigating, taking corrective action, and otherwise addressing those behaviors. The

**TABLE 4-1** Research Integrity Policy Time Line

| Year | U.S. Policy Changes | Important Contemporary Events |
|---|---|---|
| Post–World War II | | Experiments on prisoners by the Nazis and Japanese military during WWII uncovered. |
| 1966 | Animal Welfare Act (P.L. 89-544) signed into law, providing for USDA oversight and regulation of facilities performing research on laboratory animals. | |
| 1972 | | Tuskegee syphilis experiment becomes public. |
| 1974 | Department of Health, Education, and Welfare (DHEW) raises the National Institutes of Health's (NIH's) Policies for the Protection of Human Subjects (issued in 1966) to regulatory status. The regulations established the institutional review board as one mechanism through which human subjects would be protected.<br>National Research Act (P.L. 93-348) signed into law, creating the National Commission for the Protection for Human Subjects of Biomedical and Behavioral Research. | |

*continued*

**TABLE 4-1** Continued

| Year | U.S. Policy Changes | Important Contemporary Events |
|---|---|---|
| Mid-1970s Early 1980s | | Several cases of research misconduct are uncovered and widely publicized, including Summerline, Soman, and Darsee cases. |
| 1979 | | Belmont Report released. |
| 1981 | The Department of Health and Human Services (HHS, formerly DHEW) and the Food and Drug Administration revise human subjects protection regulations based on work by the National Commission for the Protection for Human Subjects of Biomedical and Behavioral Research (the Belmont Report). HHS regulations are contained in Title 45, Part 46 of the *Code of Federal Regulations*. These regulations were revised in 1983 and 1991. | Investigations and Oversight Subcommittee of the House Science and Technology Committee, chaired by Rep. Albert Gore, Jr., holds hearings on fraud in biomedical research. |
| 1985 | Health Research Extension Act (P.L. 99-158) signed into law. Under one provision, HHS requires Public Health Service (PHS) funding applicant or awardee institutions to establish "an administrative process to review reports of scientific fraud" and "report to the Secretary any investigation of alleged scientific fraud which appears substantial." NIH also established "a process for receiving and responding to reports from institutions." This legislation complemented existing authority under which the PHS pursued research misconduct in the 1970s and early 1980s. Guidelines were published in the NIH Guide for Grants and Contracts in July 1986; the Final Rule, "Responsibilities of Awardee and Applicant Institutions for Dealing with and Reporting Possible Misconduct in Science," was published in the *Federal Register* on August 8, 1989, and codified as 42 CFR Part 50, Subpart A. | |

**TABLE 4-1** Continued

| Year | U.S. Policy Changes | Important Contemporary Events |
|---|---|---|
| Mid- to late 1980s | | High-profile investigations of research misconduct allegations made against Robert C. Gallo and Thereza Imanishi-Kari receive significant media and congressional attention. |
| 1987 | National Science Foundation (NSF) establishes procedures for investigating scientific misconduct (*Federal Register*, Vol. 52, pp. 24486 ff, July 1, 1987). | |
| 1988 | | First edition of the NAS-NAE-IOM educational guide *On Being a Scientist* is published. |
| 1989 | PHS creates the Office of Scientific Integrity (OSI) in the Office of the Director, NIH, and the Office of Scientific Integrity Review (OSIR) in the Office of the Assistant Secretary for Health (OASH).<br>NSF creates the Office of Inspector General, which assumes responsibility for investigating scientific misconduct. | |
| 1990 | NIH mandates responsible conduct of research training under certain training grants. | |
| 1991 | Adoption of the Federal Policy for the Protection of Human Subjects ("Common Rule") by 16 federal agencies that conduct, support, or otherwise regulate human subjects research; the FDA also adopted certain of its provisions. | |
| 1992 | OSI and OSIR are consolidated into the Office of Research Integrity (ORI). HHS also establishes a mechanism for scientists formally charged with research misconduct to receive a hearing before the Research Integrity Adjudications Panel of the Departmental Appeals Board, HHS. | *Responsible Science* is published. |
| 1994 | | Ryan Commission report is released. |

*continued*

**TABLE 4-1** Continued

| Year | U.S. Policy Changes | Important Contemporary Events |
|---|---|---|
| 1990s and 2000s | ORI extramural program supporting research and education efforts in RCR develops and grows. | |
| 1999 | Data Access Act requires that data from federally funded research be made available to requesting parties under Freedom of Information Act procedures. | |
| 2000 | Federal Policy on Research Misconduct becomes effective, establishing a common definition of research misconduct across the federal government. | |
| Early and mid-2000s | | Schön case, Hwang case, growing international interest, series of international reports. |
| 2007 | America COMPETES Act signed into law. Includes provision that applicants for NSF funding provide responsible conduct of research training to students and postdoctoral fellows participating in research. | |
| 2009 | OSTP launches federal scientific integrity activity. | |

SOURCES: ORI, 2011; OSTP, 2000.

remainder of this chapter reviews concepts and definitions of behaviors that violate the values of research, the evolution of definitions underlying U.S. federal policies, and alternatives that are used by some U.S. institutions as well as by governments and research institutions outside the United States. Rationales for different approaches are explored, and this committee's recommended framework is presented and explained.

Some issues affecting the advantages and disadvantages of alternative approaches only become clear when considering how concepts and definitions related to violations of research integrity are actually understood and utilized in specific contexts, such as institutional investigations of alleged misconduct that are overseen by federal agencies. Issues arising from implementation of these concepts and definitions are covered in Chapter 7.

In order to develop policies and implementing mechanisms that define how and under what circumstances research institutions are to be answerable to the federal government for the research-related behaviors of their employees, it is

necessary for those behaviors to be identified. It is in this context that the definitions of research misconduct and other terms have policy implications. These concepts and definitions also have a broader significance to the research enterprise and its stakeholders, since fostering high-quality research that advances knowledge requires identifying and preventing behaviors that violate the values of research (IAC-IAP, 2012).

The 1992 report *Responsible Science* put forward a framework of terms to describe and categorize behaviors that depart from scientific integrity (NAS-NAE-IOM, 1992). This framework was developed around the terms *misconduct in science*, *questionable research practices*, and *other misconduct*. One of the tasks of this committee was to examine this framework and make recommendations about whether and how it should be updated. The goal is to describe a framework of terms and definitions that is appropriate for today's environment and that advances efforts to foster research integrity.

The sources or causes of actions that violate the values of research suggest different potential responses or approaches to preventing and addressing them. If the action arises from ignorance, education and mentoring may be the most appropriate responses. If the action arises from perverse incentives in the research enterprise, the removal or mitigation of those incentives may be warranted. If the action is criminal or violates the requirements of employment contracts or research grants, then appropriate penalties or other corrective actions would apply.

However, human actions often cannot be neatly ascribed to a single one of these causes. Rather, a given action can be multiply determined and therefore call for a multifaceted response. Furthermore, the causes of research misconduct and other actions that violate the values of research generally do not all lie within the individual. The social and institutional context of research, ranging from the atmosphere within a given research group to the national governance of research systems, creates incentives and disincentives for particular actions. These issues are explored in more detail in Chapter 6.

## RESEARCH MISCONDUCT

Developing a workable definition of research misconduct requires grappling with several issues. First, actions covered by the definition should represent significant departures from research values and related norms, whether these are field-specific or more global, and also be committed with the intent to mislead or deceive.

In addition, the definition of research misconduct should have clear and logically supportable boundaries. The actions included should be distinguished from transgressions that may occur on the part of researchers, and perhaps in the context of doing research, but which are better addressed by other frameworks. This will partly depend on what those other frameworks are, meaning that a definition of research misconduct appropriate in a given country might not be

appropriate elsewhere. For example, while the United States has separate policies and regulations for dealing with accusations of fabrication of data, protecting human research subjects, and ensuring humane treatment of laboratory animals, in some countries these issues are covered by a unified regulatory framework.

Also, as will be discussed further below, research institutions themselves may choose to adopt definitions of research misconduct for the purposes of their own internal management and employment policies that are broader than the definition adopted by the federal government. In the discussion below, the appropriateness or suitability of research misconduct definitions is considered primarily from the standpoint of U.S. federal policy.

The 1992 *Responsible Science* report defined misconduct in science as "fabrication, falsification, or plagiarism in proposing, performing, or reporting research" (NAS-NAE-IOM, 1992). It added that misconduct in science does not include errors of judgment; errors in the recording, selection, or analysis of data; differences in opinions involving the interpretation of data; or misconduct unrelated to the research process. Further, failure in scientific research is to be expected, since exploration entails risks. Projects or studies that fall short of hopes and expectations are not a sufficient basis for identifying misconduct.

Since 1992 the definition of misconduct in science as fabrication, falsification, or plagiarism (FFP) has become a central feature of U.S. institutional and governmental approaches to addressing breaches of scientific integrity. In 2000 the term *research misconduct* was adopted by the Office of Science and Technology Policy (OSTP) in the Executive Office of the President as part of its Federal Policy on Research Misconduct and was defined as FFP:

> I. Research Misconduct Defined
>
> Research misconduct is defined as fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results.
>
> Fabrication is making up data or results and recording or reporting them.
>
> Falsification is manipulating research materials, equipment, or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
>
> Plagiarism is the appropriation of another person's ideas, processes, results, or words without giving appropriate credit.
>
> Research misconduct does not include honest error or differences of opinion. (OSTP, 2000)

### Alternative Definitions and Non-FFP Elements

The adoption of FFP as the definition of research misconduct by OSTP came about as part of a lengthy, contentious process. Alternative definitions were developed, considered, and debated over a period of years. At the same time and

up until today, other countries have confronted similar issues and have reached a variety of conclusions. Exploring these approaches is useful in understanding the relative advantages of the FFP-only definition of research misconduct and possible alternatives.

It is noteworthy that all of the alternative definitions of research misconduct that the committee is aware of—past or present, recommended or implemented—include fabrication, falsification, and plagiarism. The differences all emerge from the question of whether other behaviors should be included as well.

### Other Serious Deviations

Prior to the adoption of the unified federal definition of research misconduct in 2000, the U.S. Public Health Service (which oversees research supported and performed by the National Institutes of Health) defined misconduct in science as "falsification, fabrication, plagiarism, or other practices that seriously deviate from those that are commonly accepted within the research enterprise for proposing, conducting, or reporting research" (Rennie and Gunsalus, 1993). The definition specified that misconduct "does not include honest error or honest difference in interpretations or judgments of data" (Price, 2013). The National Science Foundation's definition included FFP and "other serious deviations from accepted practices in proposing, carrying out, or reporting research results from activities funded by NSF" (Price, 2013). NSF's definition also included "retaliation of any kind against a person who reported or provided information about suspected or alleged misconduct and who has not acted in bad faith."

Both the PHS and NSF definitions allowed room to consider offenses other than FFP as research misconduct. Much of the research enterprise, including research universities and the associations representing them, opposed the inclusion of elements other than FFP in federal definitions, particularly the "other serious deviations" clause. For example, *Responsible Science* states that "the vagueness of this category has led to confusion about which actions constitute misconduct in science" (NAS-NAE-IOM, 1992). Concerns have also been raised that the clause would open the door to penalizing innovative approaches to research that could potentially yield significant advances.

A concrete illustration of the disagreement over "other serious deviations" arose when the Office of the Inspector General at the National Science Foundation (NSF-OIG) used the clause to launch a misconduct investigation against an investigator who "was accused of a range of coercive sexual offenses against various female undergraduate students and teaching assistants, up to and including rape" while on research trips to foreign countries led by the investigator (Buzzelli, 1993). While Office of Inspector General officials asserted that the case supported the need for the "other serious deviations" clause, one prominent scientist argued that the case represented "a preposterous and appalling application of the definition of scientific misconduct" (Schachman, 1993).

The "other serious deviations" clause remained in the two primary federal research misconduct definitions for a number of years following this case. During that time, there do not appear to have been additional cases in which its application was controversial, or any evidence that innovative research approaches were discouraged as a result, suggesting that there is cause to be skeptical about some of the arguments made against the clause. At the same time, it is not clear that the "other serious deviations" clause has been particularly missed in the years since. In the discussion below and in Chapter 7, the specific elements that might be covered by the "other serious deviations" clause are explored in order to see whether there are research behaviors that might not be adequately investigated or subject to corrective action under current policies, and if so, whether changing the federal research misconduct policy is the best way to accomplish this. Denmark's experience with the Lomborg case and its aftermath, in which a controversial finding of "scientific dishonesty" was later overturned (discussed later in this chapter), serves as an additional cautionary example of what can occur when governments and institutions utilize a broad, nonspecific definition of research misconduct.

On the basis of current knowledge, it appears that the "other serious deviations" clause and similar formulations may not have the adverse impacts on research that some have feared, but they may introduce the risk that a controversial or mishandled case could lead to turmoil and a loss of credibility on the part of the institutions and agencies charged with addressing research misconduct.

### The Ryan Commission

In 1995, the Commission on Research Integrity was organized by Congress to "advise the Secretary of Health and Human Services and Congress about ways to improve the Public Health Service (PHS) response to misconduct in biomedical and behavioral research receiving PHS funding." Known as the Ryan Commission after its chairman, Harvard professor Kenneth Ryan, it released a report on misconduct in research and treatment of good-faith whistleblowers (Commission on Research Integrity, 1995).

The report articulated the interest of the federal government in the integrity of research it funded and concluded that the definition of misconduct should be based on the "fundamental principle that scientists be truthful and fair in the conduct of research and the dissemination of research results." The commission defined its driving concern as "What is in the best interest of the public and science?" Its work aimed to provide "vital guidance for personal and ethical judgments and decisions concerning the professional behavior of scientists."

The commission recommended broadening the definition of misconduct beyond FFP to encompass misappropriation, interference, and misrepresentation:

> 1. Research Misconduct
>
> Research misconduct is significant misbehavior that improperly appropriates the intellectual property or contributions of others, that intentionally impedes

> the progress of research, or that risks corrupting the scientific record or compromising the integrity of scientific practices. Such behaviors are unethical and unacceptable in proposing, conducting, or reporting research, or in reviewing the proposals or research reports of others.
>
> Examples of research misconduct include but are not limited to the following:
>
> Misappropriation: An investigator or reviewer shall not intentionally or recklessly
>
> a. plagiarize, which shall be understood to mean the presentation of the documented words or ideas of another as his or her own, without attribution appropriate for the medium of presentation; or
>
> b. make use of any information in breach of any duty of confidentiality associated with the review of any manuscript or grant application.
>
> Interference: An investigator or reviewer shall not intentionally and without authorization take or sequester or materially damage any research-related property of another, including without limitation the apparatus, reagents, biological materials, writings, data, hardware, software, or any other substance or device used or produced in the conduct of research.
>
> Misrepresentation: An investigator or reviewer shall not with intent to deceive, or in reckless disregard for the truth,
>
> a. state or present a material or significant falsehood; or
>
> b. omit a fact so that what is stated or presented as a whole states or presents a material or significant falsehood. (Commission on Research Integrity, 1995)

The commission based its recommendation to include "interference" as an element of misconduct based on testimony it received about cases where researchers sabotaged the experiments of others or absconded with vital data, arguing that existing laws against vandalism were often not adequate to address these situations. It also recommended defining other forms of "professional misconduct" as obstruction of investigations of research misconduct and repeated noncompliance with research regulations after notice. Finally, the commission made several recommendations concerning the conduct and oversight of investigations, including a "Whistleblower's Bill of Rights."

The Ryan Commission's proposed misappropriation, interference, and misrepresentation definition of research misconduct was opposed by some members of the research enterprise, including the leadership of the Federation of American Societies for Experimental Biology and the National Research Council of the National Academy of Sciences. The criticisms of the definition focused on two issues.[2] First, the definition took the form of "leading principles with examples," which was characterized as "vague and open-ended" (Alberts et al., 1996). The commission's report had itself argued that fabrication, falsification, and plagia-

[2] Chapter 7 will discuss issues raised by some of the other Ryan Commission recommendations.

rism as understood in the agency policies in effect at that time were "neither narrow nor precise" (Commission on Research Integrity, 1995). Second, regarding the examples themselves, the concern was raised that the inclusion of omitting facts as an example of misrepresentation could open the door to regarding omissions or mistakes in citation as misconduct (Glazer, 1997). While many of the commission's recommendations later were incorporated into governmental regulatory approaches, its approach to the definition of research misconduct was abandoned.

### Alternative Research Misconduct Definitions Used by U.S. Research Institutions and Private Sponsors

While U.S. research institutions must apply the federal research misconduct definition to federally supported work, they are free to adopt definitions of research misconduct that include behaviors other than FFP. A recent analysis found that more than half of 189 universities studied "had research misconduct policies that went beyond the federal standard" (Resnik et al., 2015). The most common non-FFP element was "other serious deviations," with more than 45 percent of institutions including it. Other misconduct elements adopted by at least 10 percent of institutions were "significant or material violations of regulations," "misuse of confidential information," "misconduct related to misconduct," "unethical authorship other than plagiarism," "other deception involving data manipulation," and "misappropriation of property/theft" (Resnik et al., 2015). Institutional investigations of non-FFP misconduct are not reported to federal agencies or reviewed by them. Most of the policies that went beyond FFP were adopted after 2001, and a higher proportion of institutions in the lowest quartile of research funding adopted such policies than those in the upper quartiles.

Nonfederal research sponsors may also adopt research misconduct definitions different from those of the federal government. For example, the Howard Hughes Medical Institute's policy, adopted in 2007, defines research misconduct as FFP and "any other serious deviations or significant departures from accepted and professional research practices, such as the abuse or mistreatment of human or animal research subjects" (HHMI, 2007).

## Non-U.S. Examples

Policy approaches to fostering research integrity vary widely around the world, and the same variety can be seen in how research misconduct is defined (or not defined). A recent survey of research misconduct policies around the world found that 22 of the top 40 R&D performing countries have national policies, and several more are in the process of developing a policy (Resnik et al., 2015). All of the countries that have policies include FFP in their definitions, with many including additional elements such as unethical authorship and publication practices,

other serious deviations, and violation of regulations protecting human research subjects or laboratory animals (Resnik et al., 2015). The following examples illustrate the choices other countries have made, which are relevant to the question of how U.S. definitions and policies operate in a global context.

Research Councils UK, the organization of the United Kingdom's government-funding agencies, has a lengthy and detailed definition of "unacceptable conduct":

> Unacceptable conduct includes each of the following:
>
> **Fabrication**
>
> This comprises the creation of false data or other aspects of research, including documentation and participant consent.
>
> **Falsification**
>
> This comprises the inappropriate manipulation and/or selection of data, imagery and/or consents.
>
> **Plagiarism**
>
> This comprises the misappropriation or use of others' ideas, intellectual property or work (written or otherwise), without acknowledgement or permission.
>
> **Misrepresentation, including:**
>
> - misrepresentation of data, for example suppression of relevant findings and/or data, or knowingly, recklessly or by gross negligence, presenting a flawed interpretation of data;
> - undisclosed duplication of publication, including undisclosed duplicate submission of manuscripts for publication;
> - misrepresentation of interests, including failure to declare material interests either of the researcher or of the funders of the research;
> - misrepresentation of qualifications and/or experience, including claiming or implying qualifications or experience which are not held;
> - misrepresentation of involvement, such as inappropriate claims to authorship and/or attribution of work where there has been no significant contribution, or the denial of authorship where an author has made a significant contribution.
>
> **Breach of duty of care,** whether deliberately, recklessly or by gross negligence:
>
> - disclosing improperly the identity of individuals or groups involved in research without their consent, or other breach of confidentiality;
> - placing any of those involved in research in danger, whether as subjects, participants or associated individuals, without their prior consent, and without appropriate safeguards even with consent; this includes reputational danger where that can be anticipated;

- not taking all reasonable care to ensure that the risks and dangers, the broad objectives and the sponsors of the research are known to participants or their legal representatives, to ensure appropriate informed consent is obtained properly, explicitly and transparently;
- not observing legal and reasonable ethical requirements or obligations of care for animal subjects, human organs or tissue used in research, or for the protection of the environment;
- improper conduct in peer review of research proposals or results (including manuscripts submitted for publication); this includes failure to disclose conflicts of interest; inadequate disclosure of clearly limited competence; misappropriation of the content of material; and breach of confidentiality or abuse of material provided in confidence for peer review purposes.

**Improper dealing with allegations of misconduct**

- Failing to address possible infringements including attempts to cover up misconduct or reprisals against whistle-blowers
- Failing to deal appropriately with malicious allegations, which should be handled formally as breaches of good conduct. (RCUK, 2013)

Another example is Denmark, whose approach has evolved over time. The first Danish Committees on Scientific Dishonesty (DCSD) was established by the Danish Medical Research Council in 1992, with additional committees being added in 1998 so as to cover all of science (Resnik and Master, 2013). At first, the DCSD employed a broad definition of scientific dishonesty based on "actions or omissions in research which give rise to falsification or distortion of the scientific message or gross misrepresentation of a person's involvement in the research," (DCSD, 2015) with nine specific elements, including FFP, as well as "consciously distorted reproduction of others' results" and "inappropriate credit as the author or authors" (DCSD, 2002).

However, in 2003, the Danish Committees on Scientific Dishonesty investigated allegations of scientific dishonesty made against Bjørn Lomborg, whose book *The Skeptical Environmentalist* challenged the view that global environmental problems are worsening. Its finding that Lomborg had committed scientific dishonesty was controversial and was ultimately overturned by Denmark's Ministry of Science, Technology and Innovation, which cited insufficient evidence and arguments and an overly broad definition of scientific dishonesty (Resnik and Master, 2013). Several years later, Denmark's definition of scientific dishonesty was narrowed to (DCSD, 2014):

> The term "scientific dishonesty" (research misconduct) is defined as: falsification, fabrication, plagiarism and other serious violations of good scientific practice committed intentionally or due to gross negligence during the planning, implementation or reporting of research results.

There have been several international efforts to foster research integrity at the regional or global levels. For example, the European Code of Conduct for Research Integrity puts forward a definition that includes FFP as well as:

> *failure to meet clear ethical and legal requirements* such as misrepresentation of interests, breach of confidentiality, lack of informed consent and abuse of research subjects or materials. Misconduct also includes *improper dealing* with infringements, such as attempts to cover up misconduct and reprisals on whistle-blowers. (ESF-ALLEA, 2011)

### Intent

In finding that a researcher has committed misconduct, intention plays a critical role. Fabrication and falsification generally are associated with an intention to deceive. If a researcher produces incorrect results out of negligence or carelessness, the behavior is typically criticized but would not be considered misconduct, since there was no conscious deception. Likewise, plagiarism is often intentional but can also result from sloppy work practices that could be characterized as "reckless." In addition to stipulating that research misconduct does not include "honest error," the federal research misconduct policy includes the provision that the behavior must be "committed intentionally, or knowingly, or recklessly" in order for a finding of misconduct to be warranted (OSTP, 2000).

Dresser (1993) has pointed out that terms such as "intentional" and "fraudulent" are too broad and poorly defined to be useful in determining the culpability of researchers and in establishing penalties and other corrective steps for a given action. She pointed instead to the 1962 publication of the Model Penal Code, which sought to replace "eighty or so" culpability terms previously found in state and federal criminal codes with four culpable mental state provisions (American Law Institute, 1985). Individuals act "purposely" if their "conscious object" is to engage in proscribed conduct. They act "knowingly" if they are aware of a high probability that they are engaging in such conduct. They act "recklessly" if they are aware of and "consciously disregard" a substantial risk that they are engaging in prohibited conduct. And they act "negligently" if they should be aware of a substantial risk that they are engaging in prohibited conduct. The first three terms are "subjective" culpability in which an individual has some level of personal awareness of engaging in prohibited behavior.

Distinguishing "honest error" from deception can be very difficult, yet it is important for those charged with investigating an allegation to try to do so. A classic example that illustrates this is the "cold fusion" episode of 1989 involving Martin Fleischmann and B. Stanley Pons of the University of Utah (Goodstein, 2010). While that case involved research behavior that fell far short of good research practices, many observers and experts believe that it did not rise to the level of misconduct. The Fleischmann-Pons case also featured institutional and researcher choices about pursuing press conference science and secrecy to protect

intellectual property instead of publication that remain controversial to this day. Even in the most egregious cases, a researcher may claim extenuating circumstances, negligence, or error rather than admitting culpability. Furthermore, the researcher engaging in the behavior may choose not to examine the motivations behind those acts so as to reduce personal accountability. In such cases, it can be difficult to establish culpability for a given behavior.

The intent to deceive is often difficult to prove. Proof almost always relies on circumstantial evidence, which can, however, include an analysis of the behavior of the person accused of misconduct. One commonly accepted principle, adopted by the Ryan Commission, is that the intent to deceive may be inferred from a person's acting in reckless disregard for the truth (Commission on Research Integrity, 1995). Providing guidance of this sort for misconduct investigative committees would likely be valuable going forward, given that it is often difficult to establish intent.

## Implications of Retaining FFP as the Federal Misconduct Definition and Possible Changes

The above review of the debate over the U.S. research misconduct definition and alternatives past and present reveals examples of non-FFP behaviors that could be included in an amended federal research misconduct definition. Whether they should be or not depends on whether the behavior is adequately addressed under current policies related to research misconduct and other areas and, if not, whether the behavior would be addressed most effectively by including it in the federal research misconduct definition versus other options. For example, some behaviors that are included in non-U.S. definitions of research misconduct—such as violating the rights of human research subjects—are already addressed by a well-developed set of regulations and institutions in the United States (see the discussion of "other misconduct" below). Therefore, they will not be considered further in this context. Other behaviors such as sabotaging the experiments of others or retaliating against good-faith whistleblowers are worth examining in light of how the federal policy on research misconduct is actually operating within institutions and with regard to agency oversight. These issues will be discussed in Chapter 7.

In the meantime, it is worth considering an issue that the committee spent considerable time discussing, that of authorship misrepresentation that might not be clearly included in OSTP's definition of plagiarism. A footnote in the 1992 report *Responsible Science* states that "it is possible that some extreme cases of noncontributing authorship may be regarded as misconduct because they constitute a form of falsification" (NAS-NAE-IOM, 1992). *Responsible Science* also noted that in 1989 a Public Health Service annual report of its activities to address research misconduct included several abuses of authorship in examples of misconduct, such as "preparation and publication of a book chapter listing

co-authors who were unaware of being named as co-authors," and "engaging in inappropriate authorship practices on a publication and failure to acknowledge that data used in a grant application were developed by another scientist." It should be noted that this formulation predated the 2000 federal policy on research misconduct and could have included cases considered under the "other serious deviations" provision.

As in the cases of whistleblower retaliation and sabotage, evaluating whether changes in federal policy should be made to better address authorship abuses involves considering the scale of the problem and weighing the advantages and disadvantages of policy changes against other alternatives. This will be covered in Chapter 7.

## DETRIMENTAL RESEARCH PRACTICES

The 1992 *Responsible Science* report identified an additional set of actions "that violate traditional values of the research enterprise and that may be detrimental to the research process," but for which "there is at present neither broad agreement as to the seriousness of these actions nor any consensus on standards for behavior in such matters." As examples of these actions, it cited

> failing to retain significant research data for a reasonable period, maintaining inadequate research records, conferring or requesting authorship on the basis of a specialized service or contribution that is not significantly related to the research reported in the paper, refusing to give peers reasonable access to unique research materials or data that support published papers, using inappropriate statistical or other methods of measurement to enhance the significance of research findings, and misrepresenting speculations as fact or releasing preliminary research results, especially in the public media, without providing sufficient data to allow peers to judge the validity of the results or to reproduce the experiments.

Many of the actions the 1992 panel identified as questionable research practices (often labeled QRPs) have gained less institutional consensus, and consequently there is less agreement on policies and incentives to address them. However, this panel has identified some of these practices as not questionable at all but as clear violations of the fundamental tenets of research. As will be covered in detail in Chapter 5, the past several decades of experience have clarified the damage that these practices are wreaking on the research enterprise, which might surpass the damage that research misconduct causes. Codes of responsible conduct of research in other countries include some of these practices in definitions of research misconduct that are broader than in the United States.

Also, it is important to remember that *Responsible Science* and other analyses of its time focused on the actions of individual researchers, and that their concepts and definitions were framed accordingly. In light of several decades

of subsequent experience and the massive changes in the scientific landscape detailed in Chapter 3, it is clear that the organizations that make up the research enterprise, such as research institutions, research sponsors, and journals, may also engage in behaviors that damage research integrity. It is just as necessary to identify and actively discourage these organizational actions and incentives as it is to better address individual behaviors.

This committee believes that many of the practices that up to now have been considered questionable research practices, as well as damaging behaviors by research institutions, sponsors, or journals, should be considered *detrimental research practices (DRPs)*. Researchers, research institutions, research sponsors, journals, and societies should discourage and in some cases take corrective actions in response to DRPs.

Rather than develop a definitive list and specific corrective actions, the committee seeks to catalyze discussion within the research enterprise on what can be done to more actively discourage DRPs than what has been done up to now. Indeed, the committee's primary recommended response to DRPs is for all participants in the research enterprise to seek to significantly improve practices. How this may be done is covered in detail in Chapter 9.

These are examples of DRPs that the committee has considered and agrees on:

- Detrimental authorship practices that may not be considered misconduct, such as honorary authorship, demanding authorship in return for access to previously collected data or materials, or denying authorship to those who deserve to be designated as authors;
- Not retaining or making data, code, or other information/materials underlying research results available as specified in institutional or sponsor policies, or standard practices in the field;
- Neglectful or exploitative supervision in research;
- Misleading statistical analysis that falls short of falsification;
- Inadequate institutional policies, procedures, or capacity to foster research integrity and address research misconduct allegations, and deficient implementation of policies and procedures; and
- Abusive or irresponsible publication practices by journal editors and peer reviewers.

Further discussion of DRPs, how and why they are harmful, and how they should be discouraged are topics explored in Chapter 5.

## OTHER MISCONDUCT

In addition to research misconduct and questionable research practices, *Responsible Science* identified a category of unacceptable behaviors that the panel termed *other misconduct*. These behaviors are not unique to the conduct of

research even when they occur in a research environment. Such behaviors include "sexual and other forms of harassment of individuals; misuse of funds; gross negligence by persons in their professional activities; vandalism, including tampering with research experiments or instrumentation; and violations of government research regulations, such as those dealing with radioactive materials, recombinant DNA research, and the use of human or animal subjects."

Because such actions are not unique to the research process, they do not constitute research misconduct, the panel said. They should, therefore, be addressed in other ways, such as the legal system, employment actions, or other mechanisms that address violations of professional standards. However, the panel added that some forms of other misconduct are directly associated with research misconduct, including "cover-ups of misconduct in science, reprisals against whistle-blowers, malicious allegations of misconduct in science, and violations of due process protections in handling complaints of misconduct in science." As a result, these forms of other misconduct "may require action and special administrative procedures" (NAS-NAE-IOM, 1992).

As discussed above, whistleblower retaliation and tampering/sabotage will be explored further in Chapter 7. Otherwise, this committee agrees that the category of *other misconduct* should remain as it was recommended in *Responsible Science*.

# 5

# Incidence and Consequences

*Science has the potential to address some of the most important problems in society and for that to happen, scientists have to be trusted by society and they have to be able to trust each others' work. If we are seen as just another special interest group that are doing whatever it takes to advance our careers and that the work is not necessarily reliable, it's tremendously damaging for all of society because we need to be able to rely on science.*

—Ferric Fang quoted by Jha (2012)

***Synopsis:*** *Research misconduct and detrimental research practices constitute serious threats to science in the United States and around the world. The incidence of research misconduct is tracked by official statistics, survey results, and analysis of retractions, and all of these indicators have shown increases over time. However, as there are no definitive data, it is difficult to say precisely what the incidence of misconduct is per grant or per paper and to determine trends. It* is *possible to say that while research misconduct is unusual, it is not rare. A variety of detrimental research practices appear to be tolerated, at least in the fields and disciplines that have been studied. Both research misconduct and detrimental research practices impose significant costs on the research enterprise. Particular cases of misconduct have also negatively affected society at large. The phenomenon of irreproducibility, which has attracted increasing attention during the course of this study, illustrates the negative impacts of detrimental research practices (DRPs), although this is a complex phenomenon and specifying the role of DRPs in irreproducibility will require additional research. Examining specific cases shows that tolerance for DRPs at the level of laboratories, institutions, sponsors, and journals enables misconduct and leads to delays in uncovering it. In addition, some DRPs are committed either directly or through inadequate practices by research institutions and journals, not just by individual researchers and research groups.*

## THE INCIDENCE OF RESEARCH MISCONDUCT AND DETRIMENTAL RESEARCH PRACTICES

The *Responsible Science* report (NAS-NAE-IOM, 1992) found that "existing data are inadequate to draw accurate conclusions about the incidence of mis-

conduct in science or about questionable research practices." The report pointed out that "the number of confirmed cases of misconduct is low compared to the level of research activity in the United States," but that there might be significant underreporting, and that "every case of misconduct in science is serious and requires attention."

In recent years, a regular flow of high-profile cases of fabrication, falsification, and plagiarism (FFP) has been covered in the media. These have come from countries around the world, and they have been notable due to the prominence of the researchers involved, the importance of the work shown to be false or unreliable, the scale of the transgression in terms of, say, the number of papers to be retracted, or some combination of these factors. A particular trend has been the emergence of "serial misconduct"—cases of careers built on fabrication involving up to a hundred or more publications. A few examples taken from the past few years:

- In 2012, Harvard psychologist Marc Hauser, who gained prominence for his groundbreaking work on the origins of cognition and morality, was found by the Office of Research Integrity (ORI) of the Department of Health and Human Services (HHS) to have falsified and fabricated data and methods in six federally funded studies (Carpenter, 2012).
- In 2012 the University of Connecticut found that cardiovascular researcher Dipak Das fabricated or falsified data 145 times in his work on resveratrol (*Science*, 2012).
- As of September 2012, 28 papers authored by Korean biochemist Hyung-In Moon of Dong-A University in South Korea had been retracted as a result of suspicions that he supplied reviewer suggestions to journals with e-mail addresses that actually went to him (Fischman, 2012).
- In 2012 the Japanese Society of Anesthesiologists released a report on the work of Toho University of Medicine faculty member Yoshitaka Fujii, concluding that he had fabricated data in 172 papers (JSA, 2012).

Over the past several decades, as federal agencies and research institutions have had to address research misconduct more frequently and institute formal policies, more information has become available about the incidence and significance of research misconduct. Information on the incidence of research misconduct, defined as FFP, is available in the reports of the National Science Foundation's Office of Inspector General (NSF-OIG) and ORI.

In the case of NSF-OIG, misconduct findings have undergone a notable increase in recent years. In its semiannual reports to Congress, NSF reported just 1 finding of misconduct in 2003, 2 in 2004, and 6 in 2005, compared with 17 findings in 2012, 14 in 2013, and 22 in 2014. A rate of 16 findings per year represents less than two hundredths of a percent of the new awards NSF makes. A large proportion of NSF's research misconduct findings are for plagiarism

(e.g., 18 of 22 in 2014). The number of research misconduct allegations made to NSF-OIG annually has more than tripled over the past decade (Mervis, 2013).

Research misconduct findings by ORI have shown less of an upward trend in the past decade, with 12 findings in 2003, 8 in both 2004 and 2005, 14 in 2012, 12 in 2013, and 13 in 2014. The majority of HHS's research misconduct findings are for fabrication or falsification. As with NSF-OIG, the number of allegations made to ORI has increased significantly, going from 240 in 2011 to 423 in 2012 (ORI, 2013). Just as statistics on arrests or convictions will tend to undercount the number of crimes actually committed, the statistics on research misconduct findings will tend to undercount the actual incidence (Steneck, 2006).

In addition to these official statistics, a number of surveys of researchers regarding their practices have been undertaken in recent years. For example, a survey of research psychologists found that between a quarter and a half of the respondents admitted to having engaged in such practices as "failing to report all of a study's conditions," "selectively reporting studies that 'worked,'" and "reporting an unexpected finding as having been predicted from the start" (John et al., 2012). In an earlier survey of scientists funded by the National Institutes of Health (NIH), less than 1 percent of respondents self-reported engaging in falsification of data and less than 2 percent admitted to plagiarism, but more than 10 percent admitted to engaging in practices such as "inappropriately assigning authorship credit" or "withholding details of methodology or results in papers or proposals" (Martinson et al., 2005).

Similarly, a meta-analysis of researcher surveys indicates that the incidence of FFP is somewhat higher than the official statistics indicate, with about 2 percent of researchers admitting to fabricating or falsifying data at least once, and more than 14 percent aware of colleagues having done so (Fanelli, 2009). Survey reports on misconduct by colleagues might be inflated by multiple researchers reporting the same incidents; one of the surveys attempted to avoid this by not including more than one researcher from a given department and found that 7.4 percent of respondents had observed misconduct committed by colleagues (Titus et al., 2008). At the same time, the narrower group of respondents would not be expected to know about all cases of misconduct among colleagues, making this a conservative estimate. The same meta-analysis showed that actions discussed in Chapter 4 as examples of detrimental research practices (DRPs) are relatively common. A survey on violations of research regulations—including human subjects protection violations as well as research misconduct—was sent to all comprehensive doctoral institutions and medical schools in the United States and yielded responses from 66 percent (DuBois et al., 2013a). The results reinforce the federal agency data cited above showing a significant rise in allegations—96 percent of the responding institutions had undertaken an investigation in the preceding year, with the modal number being 3 to 5 per year.

Determining the incidence of plagiarism and related trends faces some particular barriers. The difficulty in defining plagiarism continues to be an obstacle.

While plagiarism detection software has recently grown in popularity, text matches are not necessarily plagiarized (Wager, 2014). Text matches may occur for a variety of reasons, including copublication, legal republication, common phrases, and multiple versions of a publication (Wager, 2014). However, there are indications that the overall level of plagiarism in legitimate biomedical journals peaked at some point in the last decade and has been declining since as the use of plagiarism detection software by journals has become widespread (Reich, 2010a). Despite the likely decline in incidences, differences persist between journals in how they respond to plagiarism allegations (Long et al., 2009). The appearance of a large number of journals that appear to have little concern about publishing copied or duplicated work—many of which operate under an author-pays, open-access business model—has created a new channel for papers to be plagiarized (Grens, 2013a).

Other recent research has examined retractions of scientific articles in journals (Fang et al., 2012; Grieneisen and Zhang, 2012; Steen et al., 2013). Articles may be retracted for a number of reasons, including unintentional errors on the part of authors or publishers as well as research misconduct. One recent analysis that focused on articles contained in the PubMed database found that more than two-thirds of retractions were due to misconduct defined as FFP (Fang et al., 2012). Another analysis that examined retractions of articles in a variety of databases that collectively covered all disciplines between 1980 and 2011 found that 17 percent of the 3,631 retractions in which a cause was identified were due to data fabrication or falsification, and 22 percent were due to plagiarism (Grieneisen and Zhang, 2012). This research also found that there are more retractions in certain disciplines than would be expected based on their representation in the overall research literature (e.g., biomedicine, chemistry, and life sciences) and that other disciplines are underrepresented in terms of retractions (e.g., engineering, physics, and social sciences) (Grieneisen and Zhang, 2012).

These analyses have also found a sharp increase in the number of retractions over time, particularly over the past decade or so. Although the increase in the number of articles published annually is a contributing factor, the rate of retraction is also increasing. For example, an analysis of papers in the PubMed database found that the number of retractions has increased tenfold in recent years, while the total number of papers has only increased by 44 percent (Van Noorden, 2011). As with other statistics cited here, there are reasons to be cautious about using the number and rate of retractions as proxies for the incidence of misconduct or error. One analysis suggests that both the barriers to publishing flawed work and to retracting articles have been lowered over time (Steen et al., 2013). Retraction rates, particularly at the country and disciplinary level, can be skewed by the serial misconduct cases mentioned above, where a researcher has fabricated or falsified data underlying tens of articles (Grieneisen and Zhang, 2012). On the one hand, retracted papers still represent only a small proportion of the overall literature, and formal retractions have only gradually become a standard practice

in recent decades. On the other hand, some evidence suggests that many fraudulent papers are never retracted (Couzin and Unger, 2006) and all of the admittedly imperfect proxy measures for the incidence of research misconduct have displayed significant increases in recent years.

## COSTS AND CONSEQUENCES

Research misconduct and DRPs constitute failures to uphold the values of science. Even if they had no wider consequences, it would be vital to prevent and address them. However, a variety of costs and consequences can be conceptualized even if they are difficult to quantify or measure precisely. The costs of research misconduct and DRPs can be broken down into (1) damage to the individuals, (2) reputational costs to the employer of the transgressor and the journal that published the work, (3) direct financial costs, (4) broader social costs, and (5) opportunity costs associated with categories 1 through 4. Figure 5-1 illustrates these costs.

Examples of the many individual costs of research misconduct and DRPs are wasted efforts of researchers who trusted a fabricated paper and did work to build on it, damage done to innocent collaborators (including graduate students and

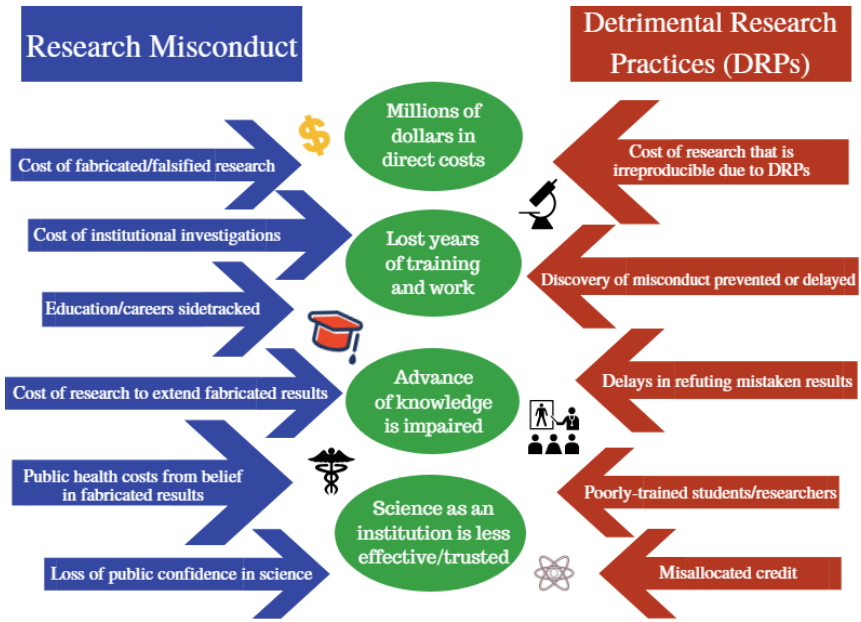


FIGURE 5-1 Costs and consequences of research misconduct and detrimental research practices (DRPs).

postdocs experiencing career turmoil after misconduct committed by a supervisor or colleague is uncovered), time and energy associated with misconduct inquiries and investigations on the part of committee members and staff, wasted time by editors and reviewers, the damaged careers of the perpetrators themselves, and any retaliation or other negative repercussions suffered by good-faith whistleblowers or informants. One measure of the wasted efforts of later researchers is the extent to which papers based on fabricated data are cited, even if they are retracted (surprisingly often) (Neale et al., 2007). To give one example, in the 1990s the Geological Survey of India and Panjab University found that paleontologist Viswa Jit Gupta had fabricated and falsified data on fossil discoveries over more than 20 years (Jayaraman, 1994). Articles citing Gupta's work are still cited, illustrating that the task of correcting the scientific record can become a long-term undertaking.

Reputational costs include the losses in prestige experienced by research institutions employing the author of a fabricated or falsified paper and by the journals publishing it.

Direct financial costs are borne by a number of stakeholders. Costs can include the funds provided by federal or private sponsors spent on fabricated or falsified research, the expense of investigating an allegation borne by the institution, and any additional funds that the institution pays to settle civil litigation connected with the misconduct. There have been efforts to directly measure the costs of research misconduct in particular cases or groups of cases. For example, a 2014 analysis found that direct NIH funding for 149 articles retracted due to research misconduct between 1990 and 2012 totaled $58 million, far less than 1 percent of NIH's budget over that period (Stern et al., 2014). This method of analysis has limitations, since the research underlying articles is often supported by multiple sources and funding may not be cited. Funding for an additional 142 articles retracted due to misconduct over that period could not be tracked completely (Stern et al., 2014). Extrapolating the average grant amount associated with the 43 retracted articles supported only by cited NIH grants ($425,073) to the entire set of 291 retracted articles would yield a total of $123.7 million, "which might be considered an estimate of the total NIH funds directly spent on known biomedical research retracted due to misconduct over the past 20 years" (Stern et al., 2014). Adding up all the grants that contributed in any way to papers retracted due to misconduct over those 20 years, which the authors point out would overstate the costs of misconduct, totals $1.67 billion in actual funds and $2.32 billion in 2012 dollars (Stern et al., 2014). This analysis only looks at cases where an investigation has been completed and findings of misconduct have been made.

In another example of an effort to estimate the direct costs of funding for research that is fabricated or falsified, a report by the U.S. Department of Interior's Office of Inspector General on a scientific integrity incident at a U.S. Geological Survey laboratory states that research and assessment projects totaling $108 mil-

lion in funding between 2008 and 2014 were affected by erroneous data produced by one individual researcher (DOI-OIG, 2016). Further analysis would be needed to determine the specifics of how these projects were affected and the actual costs of wasted effort and any work that had to be redone.

The reproducibility problem will be discussed in more detail below, but if the conclusions being drawn by some researchers in this area are anywhere close to correct, billions of dollars of public and private support for research might not be producing reliable knowledge (Ioannidis, 2005). The cost of research misconduct investigations should also be considered. One analysis has estimated the typical cost of an investigation for an institution to be $520,000 (Michalek et al., 2010). If this amount is extrapolated as the average for the 217 investigations reported to ORI during a year prior to the analysis, it would imply that an annual total of $110 million is spent by institutions on misconduct investigations involving HHS-funded research.

One area where the broader social costs of research misconduct are apparent from specific historical cases is in biomedical research. For example, research characterized by misconduct and DRPs funded by the tobacco companies very likely delayed the issuance of a warning on smoking and health by the U.S. Surgeon General. One expert estimates that if the warning had come out in 1959 rather than 1964, gains from an earlier decline in smoking in the form of increased life expectancy would have totaled $27 billion (Gardner, 2006). In another case, the fraudulent work of anesthesiologist Scott Reuben of Baystate Medical Center in Massachusetts played a large role in shaping treatments in that area over the years (White et al., 2009). This raises the possibility that deaths and other adverse events occurred due to administering treatments developed on the basis of fraudulent work, and it has necessitated researchers going back over the literature to see what findings can be salvaged and what experiments need to be redone (White et al., 2009). The case of Don Poldermans of the Erasmus Medical Center in the Netherlands has caused similar damage, although Poldermans denies having fabricated his work and the Erasmus Medical Center report has not been made public (Chopra and Eagle, 2012). The work of Poldermans and his collaborators in the area of perioperative use of beta blockers and statins informed clinical practice all over the world. The appropriate patient treatments are now highly uncertain (Bouri et al., 2013).

The case of Andrew Wakefield's finding of a possible causal link between the measles, mumps, and rubella vaccine and autism might also be considered in this context (see Appendix D). Wakefield was later removed from the United Kingdom's medical register due to professional misconduct committed while performing this research, and is alleged to have falsified data (Godlee et al., 2011; Triggle, 2010; UK GMC, 2010). The costs to society include an ongoing public controversy in multiple countries, public health costs, and even deaths due to a rise in cases of measles. It is not possible to determine the effect of Wakefield's work on decreased vaccination rates and resulting outbreaks with any precision.

See Appendix D for a more detailed write-up of this case. Cases such as these may also sow broader mistrust of researchers in society.

To the extent that fabricated papers impede drug and treatment development by leading researchers down the wrong track, they also impose financial costs on companies and public health costs on society.

## The Reproducibility Problem, Research Misconduct, and Detrimental Research Practices

Meta-analyses of research on particular research questions and even entire fields have produced new insights on the reliability of research. Apparently high rates of irreproducibility of research results in fields such as preclinical biomedical research and social psychology have been discovered and discussed over the past decade (Ioannidis, 2005; OSC, 2015). Issues related to reproducibility began to receive more general attention starting with a cover story that appeared in the *Economist* in October 2013 (*Economist*, 2013). The President's Council of Advisors on Science and Technology devoted a significant amount of its January 2014 meeting to a discussion of the issue (McNaull, 2014). The journal *Nature* has set up an archive of articles on challenges in irreproducible research (*Nature*, 2015b). The reproducibility problem has gained wide attention and recognition as a major issue in science.

Reproducibility can be conceptualized or defined in several ways, depending on the discipline or context. It is possible to replicate some work by "using precisely the same methods and materials" to independently collect data, which would require the original work to be presented in "sufficient detail to allow replication or reanalysis" (Freedman et al., 2015). For example, it should be straightforward to replicate a chemical reaction if the amounts of the chemicals to be combined and other conditions such as pressure and temperature are specified precisely. Observations of many natural phenomena cannot be replicated exactly, but precise descriptions of a given phenomenon and the analytical methods used will allow others to validate the conclusions drawn by observing and analyzing similar phenomena. Likewise, clinical trials cannot be replicated exactly, even if the same dosage of a given pharmaceutical is tested on the same number of research subjects with similar characteristics, since the population being tested is different. However, if a drug has an actual, measurable therapeutic effect across a given population of subjects, the effect should be observed when the drug is administered to a similar population. In fields where replication through the independent collection and analysis of data is difficult or impossible for cost or other reasons, a different standard for reproducibility might be used, in which data and the computer code used to analyze them are made available to others for validation (Peng, 2011).

The failure to reproduce research results may be due to a number of factors. This is a nonexhaustive list:

- One or more independent variables affecting the results were not characterized or measured in the original work;
- One or more errors were made in setting up the experiment, data collection or recording, or data analysis in the original work;
- Reporting of the experimental procedures, data obtained, analytical methods, or other aspects of the original work were incomplete;
- The data in the original work were correctly obtained and recorded, but the reported results constituted a false positive or false negative;
- Data in the original work were fabricated or falsified;
- One or more errors were made in setting up the experiment, data collection or recording, or data analysis in an effort to reproduce the work.

In the normal progress of research, a certain level of irreproducibility is to be expected. If irreproducibility is due to unknown variables, knowledge advances when these are characterized and understood through further work. A certain level of error and false positives is compatible with a healthy field. Past a certain point, efforts to eliminate all of the possible factors that cause irreproducibility would be prohibitively expensive (Freedman et al., 2015).

However, concerns have been raised in recent years as irreproducibility rates of 50 percent or more have been estimated in certain fields. This is a far higher level than what might be considered healthy and implies that a significant fraction of effort in some fields is not advancing knowledge. In clinical research, for example, the prevalence of studies with relatively few participants, the reporting of effects that are small by statistical measures, a high number of tested relationships, greater flexibility in study design, the involvement of researchers with personal financial interests, and the popularity of a topic are correlated with the incidence of false positive results. According to a widely discussed analysis, systematic biases led to false positive findings in half or more published studies (Ioannidis, 2005). In addition, false claims may continue to be cited at a high rate, even after subsequent published studies have refuted them (Rekdal, 2014; Tatsioni et al., 2007). The low quality of preclinical research has been identified as a significant factor in the high failure rate of clinical trials in oncology. A recent effort to replicate 53 landmark preclinical studies in hematology and oncology was successful for only 6 articles (Begley and Ellis, 2012).

Lack of reproducibility has also become a significant issue in psychology. A large-scale effort to replicate 100 results published in psychology journals found that the mean effect size of the replications was about half of what was reported in the original articles, and that while 97 of the original articles reported significant results, only 36 of the replications did (OSC, 2015).

Contemporary concerns about reproducibility arise in several forms (Academy of Medical Sciences et al., 2015). One relates to the well-known bias toward reporting and publishing positive results on the part of researchers—the "file drawer problem" (Rosenthal, 1979). The measured effect in a study is the combi-

nation of any real effect plus random variability. Studies where random variability augments the real effect have high formal statistical significance. They are more likely to be submitted to and accepted by journals than work where random variability diminishes the real effect and statistical significance is not achieved. This bias can create a situation in which false-positive results are overrepresented in the published literature, particularly for research with small cohorts, even when the data are correct and the effects are real. When random variability augments an effect in a primary study, it is not likely to do so to the same extent in a replication study. This is an example of the well-known statistical phenomenon of regression to the mean. In some fields, it may be difficult or impossible to quantitatively predict the size of the effect or determine the cohort size likely to produce results that are statistically significant.

Experts have argued that the incentive structures in many modern research environments exacerbate this problem (Nosek et al., 2012). High-pressure research environments, poor publication practices, and funding patterns that create perverse incentives are presumed to be contributing factors (Alberts et al., 2014). These issues are explored in more detail in Chapter 6. An extreme form of positive results bias is seen in the practice of "p-hacking," in which a dataset is searched or analyzed for a statistically significant relationship, to which a theory of causality is then attached (Academy of Medical Sciences et al., 2015). In general, caution is required when hypotheses are formulated after data have been collected; arcane hypotheses with marginal significance should be regarded with great suspicion. Statistical methods for testing multiple previously defined hypotheses at the same time with a dataset are available (Benjamini and Hochberg, 1995).

Another set of reproducibility concerns arises from flaws in study design and planning (Academy of Medical Sciences et al., 2015). An experimental design may be flawed to the point where it cannot be expected to produce reliable results, or the sample size may be too small to reliably confirm a statistically significant effect, leaving the study "underpowered" (Academy of Medical Sciences et al., 2015).

Other sources of error also figure into the discussion of reproducibility. The growing dependence of many fields of research on information technology and computational science, particularly in areas such as data analysis and simulation, is one potential source of error (Donoho et al., 2008). If data and the code used to analyze the data are not made available, the results cannot be validated through reanalysis. Another source of error that has become problematic in biomedical research is the lack of validation of certain reagents, including antibodies, cell lines, and animal models (GBSI, 2015). The widespread misidentification of cell lines is a specific example (*Nature*, 2015b).

To what extent are research misconduct and DRPs implicated in the reproducibility problem? There is still much to be learned about reproducibility, both in general and in specific fields. While results based on data fabrication and falsification would certainly be irreproducible, they would constitute only a small part of the reproducibility problem being faced in fields such as biomedical

research and psychology. Certain DRPs, such as misleading statistical analysis that falls short of falsification and the practice of p-hacking, are DRPs that are a direct cause of irreproducibility. Other DRPs, such as failing to share data and code, make replication and validation of results difficult or impossible and are therefore part of the reproducibility problem. Inattentive supervision of postdocs and other junior researchers and failure to catch obvious errors is another DRP that underlies some lack of reproducibility. In addition, tolerance for DRPs that cause or exacerbate the reproducibility problem on the part of journals, research institutions, and sponsors can make it more difficult to uncover research misconduct, as discussed below.

Regarding the costs of irreproducibility, one recent analysis puts forward an estimate, intended to be used as a starting point for debate, of $28 billion per year spent in the United States on "research that cannot be replicated" in preclinical biomedical research alone (Freedman et al., 2015). This figure is not based on a cost analysis but was created by applying a 50 percent rate of irreproducibility, around the lower bound of estimates generated by recent studies, to the total amount of preclinical biomedical research performed in the United States. The uncertainty surrounding this estimate points to the need to better quantify the costs and causes of the reproducibility problem in specific fields and across the research enterprise.

In addition to the direct financial costs, results that are irreproducible due to DRPs have some indirect costs that are similar in type to those that are incurred due to research misconduct, such as delays in rejecting and confirming key results, the time and effort of the researchers involved, and the time and effort of those seeking to build on false results. Chapter 7 will explore how DRPs can be more effectively uncovered and addressed.

### Connections Between Detrimental Research Practices and Research Misconduct

Developments in social psychology demonstrate a link between a field's tolerance for DRPs and delays in discovering significant cases of fabrication and falsification. Social psychology has received scrutiny recently due to a string of high-profile misconduct cases and doubts about the reliability of key results. In 2011, concerns were raised about the work of noted Dutch researcher Diederik Stapel, and a subsequent investigation by the three universities where he studied and worked found that he had fabricated data in 55 publications over many years (Levelt et al., 2012).

The Stapel investigation report enumerates a long list of DRPs that were used by Stapel and his coauthors and that appear to have been widely tolerated in the social psychology research culture. These include a variety of practices reflecting verification bias, such as repeating an experiment that has failed to produce the expected statistically significant result with minor changes in conditions—changes that would not be expected to affect the result—until statistically sig-

nificant results are attained, then reporting only those results. Often, incorrect or incomplete information about research procedures was provided in the publication. Statistical errors that reflected a lack of understanding of elementary statistics were common.

Perhaps the most alarming finding in the Stapel investigation report is the failure of coauthors, editors, and reviewers of leading social psychology journals and others in the field to note infeasible experiments or impossible results. Indeed, reviewers often reportedly encouraged DRPs in the service of "telling an interesting, elegant, concise and compelling story" (Levelt et al., 2012). The report concludes that "there are certain aspects of the discipline itself that should be deemed undesirable or even incorrect from the perspective of academic standards and scientific integrity."

Daniel Kahneman, who won the 2002 Nobel Memorial Prize in Economic Sciences for his work on the psychology of decision making, challenged social psychologists in a 2012 e-mail message about research on priming, the phenomenon where exposure to a stimulus increases sensitivity to a later stimulus:

> For all these reasons, right or wrong, your field is now the poster child for doubts about the integrity of psychological research. Your problem is not with the few people who have actively challenged the validity of some priming results. It is with the much larger population of colleagues who in the past accepted your surprising results as facts when they were published. These people have now attached a question mark to the field, and it is your responsibility to remove it. (Kahneman, 2012)

### Estimating a Range of Financial Costs of Research Misconduct and DRPs

From this discussion and the existing evidence, it is possible to develop a reasonable range of the estimated costs borne by the research enterprise and the broader society due to research misconduct and DRPs. For example, the analysis discussed above estimated that confirmed cases of research misconduct directly affected about one-tenth of one percent of NIH extramural funding over the 1992-2012 period, implying an annual total of about $30 million for one agency if this relationship were to continue going forward. To this, one could add the cost of supporting work that is confirmed to be falsified or fabricated work by other federal agencies and the private sector, the cost of supporting falsified or fabricated work that is never investigated by all funders, and the indirect costs of supporting research to extend this fraudulent work. The cost of institutional investigations is estimated at $110 million per year (Michalek et al., 2010). Taking these various costs into account, a total of several hundred million dollars a year would be a reasonable, conservative estimate of the direct financial costs of research misconduct.

Indirect costs such as those arising from negative public health impacts that fabricated and falsified research contribute to, discussed above, should also be

included. The historical case of the tobacco industry and the more recent case of vaccines illustrate that these costs may run into the millions or even billions of dollars over a period of years in particular cases when the public is misinformed on important health issues. The costs of years of incorrect treatment given to thousands or even millions of patients and the costs of accumulating new knowledge to develop correct treatments, as illustrated by the Reuben and Poldermans cases, should also be considered.

DRPs also impose costs on the research enterprise. The financial costs of DRPs in the form of funding for research that does not produce reliable knowledge may be even larger than the analogous costs of research misconduct. There is much still to be learned about irreproducibility in research, including the extent to which DRPs are implicated and how significant a problem it is in fields other those where it is being actively examined such as biomedical research and social psychology.

Another consideration is the international nature of costs and consequences. This discussion focuses on the costs and consequences for the United States, but the Wakefield case shows that misconduct or DRPs committed elsewhere in the world can impose significant costs on U.S. patients and communities. The reverse is true as well.

Clearly, the costs of research misconduct and DRPs are currently difficult to estimate. From the above discussion, and taking into account estimates of several categories of costs, several hundred million dollars in annual costs within the United States is a reasonable lower bound, and the total may be as high as several billion dollars.

# 6

# Understanding the Causes

*[H]ow dishonesty works . . . depends on the structure of our daily environment.*

—Dan Ariely (2012)

***Synopsis:*** *Improving understanding of why researchers commit misconduct and detrimental research practices (DRPs) is important because this understanding should inform the responses of the research enterprise and its stakeholders. For instance, if the only perpetrators of research misconduct and DRPs are a very small number of bad people engaged in self-interested deception and shortcuts, then the potentially useful responses of the research enterprise might be limited to increased vigilance in uncovering these "bad apples" and ending their research careers. To the extent there are other factors contributing to research misconduct and DRPs, such as institutional environments for research integrity or incentive structures significantly shaped by the policies and practices of journals and funding agencies, then other responses are required. Recent advances in understanding human cognition have implications for the response of the research enterprise to problems.*

*Why people engage in criminal or other pathological behavior and the conditions that encourage or discourage such behavior are issues of perennial interest in the behavioral and social sciences. Recent work provides some useful insights on these questions that are relevant to understanding why and under what conditions researchers commit misconduct and engage in detrimental research practices. Current patterns of U.S. research funding and organization are contributing to research environments with characteristics that behavioral and social sciences research suggests facilitate and encourage detrimental behavior in science, with some evidence of negative effects. More research on the causes of research misconduct and DRPs is needed in order to develop better strategies for prevention.*

## WHY IS IT IMPORTANT TO BETTER UNDERSTAND THE CAUSES OF RESEARCH MISCONDUCT?

Beliefs and assumptions about the causes of research misconduct can shape the responses of the research enterprise and its constituent stakeholders. For example, one theory that has been expressed by scientists is that misconduct is rare

and due to the "ineradicable presence of fraudsters" (James, 1995). Under this formulation, not much can be done by research institutions or others to prevent misconduct or foster integrity; the fraudulent "bad apples" can only be discovered and removed from the scientific barrel through the process of others trying and failing to replicate their work.

Indeed, current policies and practices for addressing research misconduct, described in Chapter 4 and Chapter 7, largely focus on the behavior of individuals. Specifically, federal policy defines prohibited individual behaviors as research misconduct and sets out procedures for investigating the individuals alleged to have engaged in this behavior. The policy also covers the corrective actions that might be taken against individuals in response. In the view of one expert, "The 'bad-apple' metaphor represents an old ideology, protective of science but, at the same time, perpetuating an ineffective way of dealing with research misconduct" (Redman, 2013).

Alternatively, a broader understanding that includes theories of misconduct in which individual failings interact with aspects of the immediate lab or institutional research environment—or even with larger structural conditions in research such as competition for funding or workforce imbalances—to cause a higher or lower incidence of misconduct would lead to different response strategies than those based on the bad-apple theory. Interventions directed at individual researchers that go beyond the investigation of alleged misconduct, such as better education and training or closer supervision, might be combined with efforts to improve research environments or even address structural issues.

When *Responsible Science* was released, the potential but as yet undocumented and little-understood importance of environmental factors in affecting integrity in science was acknowledged in the statement that "factors in the modern research environment contribute to misconduct in science" (NAS-NAE-IOM, 1992). A range of possible reasons were posited: (1) career and funding pressures, (2) institutional failures of oversight, (3) commercial conflicts of interest, (4) inadequate training, (5) erosion of standards of mentoring, and (6) part of a larger pattern of social deviance. *Responsible Science* specifically "made no judgment about the significance of any one factor," concluding that the alternative "bad person" and "environmental factors" hypotheses are "possibly complementary."

A similar stance is seen in *Integrity in Scientific Research: Creating an Environment That Promotes Responsible Conduct* (IOM-NRC, 2002). This report explicitly recognized the important role of the local environment—the lab, the department, the university—in shaping the behavior of scientists. Like the 1992 report before it, the 2002 Institute of Medicine–National Research Council report took an essentially agnostic stance about larger structural influences on the integrity of research practices, citing the lack of specific empirical evidence to guide policy.

If more reliable knowledge about the causes of misconduct can be attained, including the likely role of environmental factors and their interaction with the

psychology and cognitive limitations of individuals, the research enterprise and its constituent stakeholders will be able to use this knowledge to refine their approaches to preventing misconduct as well as to discovering and addressing misconduct after it has occurred. Examining the evidence on this topic bears directly on several elements of the committee's task statement, including the need to assess "the impacts on integrity of changing trends in the dynamics of the research enterprise," "the advantages and disadvantages of enhanced educational efforts," and "the appropriate roles for government agencies, research institutions and universities, and journals in promoting responsible research practices."

Choosing to stick with assumptions that are not supported by evidence as the basis for strategies to prevent and address research misconduct and detrimental research practices (DRPs) may perpetuate suboptimal responses on the part of the community, causing the negative consequences and damage resulting from misconduct that are described in Chapter 5 to be greater than they need to be.

## INSIGHTS FROM THE SOCIAL AND BEHAVIORAL SCIENCES ABOUT RESEARCH MISCONDUCT AND DETRIMENTAL RESEARCH PRACTICES

Decades of research in the social and behavioral sciences have generated important insights on why humans deviate from the behavioral norms of the groups to which they belong. Recent research has yielded insights with implications for understanding the causes of research misconduct and DRPs. While some work has been done to apply broader social and behavioral sciences insights to researchers and research environments, further efforts along these lines hold the promise of helping the research enterprise to better address research misconduct and DRPs and even reduce their incidence.

For decades, examinations of research misconduct and DRPs have been framed around concepts of deviance, explained most simply by reference to psychopathology, moral defect, or poor upbringing—in short, factors residing primarily within the individual, and typically seen as predating their involvement in research. This parallels the evolution of understanding of human conduct in other arenas, but this perspective falls short in explaining research misconduct and DRPs for several reasons: (1) the individual defects that supposedly lead to deviant behavior in science are vague and unmeasured; (2) the characteristics that supposedly define defective individuals may also be characteristics that are highly valued among eminent scientists, such as creativity, original thinking, and self-assurance; and (3) many of these traits have been observed among "scientists whose actions or ideas are controversial or inconvenient, including whistleblowers" (Gino and Ariely, 2011; Hackett, 1994). Hackett also asserts "the individualistic explanation is too convenient and too self-serving of the interests of established scientists to be accepted on faith and assertion without evidence" (Hackett, 1994).

## Evolution of Thinking on Causes of Deviant Behavior

Across a diverse range of fields and theories, a broad spectrum of causes of deviant behavior has been explored, including mental illness and moral defect, criminality, institutional failures, rational response to perverse incentive structures, and nonrational behavior arising from cognitive limitations, biases, or impaired decision making. Considerations of potentially motivating factors, as well as potentially mitigating influences, have been similarly broad, ranging from avarice to hubris to loss aversion to maladaptive coping. Additionally, views of human behavior as being largely intentional and rational have evolved to recognize the frequent presence of unintentional and nonrational elements to human behavior. Just as we now know that people tend to eat more when food is placed on larger plates, evidence is emerging that the conduct of those around us and the structure of the environment in which we work affects the integrity of choices made in performing work (Ariely, 2012; Wansink and van Ittersum, 2006). Examining these past conceptualizations can inform thinking about research misconduct and DRPs and demonstrate that this is an arena worthy of further empirical investigation.

The deviance approach that is prominent in fields such as the sociology of crime and delinquency generally holds that there are "good" people, who behave well, and there are "bad" people, who behave badly, or good people who make bad choices for reasons of personal defect or gain (Ben-Yehuda, 1986; Folger and Cropanzano, 1998; Hirschi, 1969; Matza, 1964, 1969; Sovacool, 2008). A second group of theories seen in fields such as organizational psychology, behavioral economics, and decision science is aimed primarily at understanding behavior that, while it may still be somewhat intentional, may result from biased, nonrational, and in some cases subconscious cognitive processes (Kahneman and Tversky, 1979; Thaler and Sunstein, 2008; Vaughan, 1999). Despite the diversity in viewpoints, these theoretical perspectives mainly conceive of deviance as arising from the interaction of individuals with salient aspects of their social environments.

Considerations of human behavior frequently speculate about motives for behavior. Among the motivating factors for deviance, perhaps the most commonly suggested is avarice. The simple desire for personal gain seems a natural explanation for the behaviors of self-interested individuals, not sufficiently held in check whether by self-control or threat of punishment. In fact, greed intuitively fits the "bad actor" individual defect explanation of deviant behavior, since moral defect may allow for the unhealthy expression of self-interest as greed. But while avarice may explain some deviance, it is likely too simple and convenient an explanation for most deviance in science.

Some other proposed motivating factors are thought to operate through mechanisms of human emotion or cognition. Motives consonant with recognized features of human psychology include the blockage of legitimate goals, leading to desperation, alienation, or other aversive affective states (Agnew, 1992, 2006;

Cohen, 1965; Merton, 1938). Some theories posit that deviant behavior will not result unless environmental conditions also lead to the activation of "will" and the neutralization of moral reasoning, or through the generation of negative affect (i.e., the experience of negative emotion) in individuals and their attempts at coping with that affect (Agnew, 1992, 2006; Ben-Yehuda, 1986; Matza, 1969). Other theories have suggested the importance of an intrinsic sense of justice or fairness (typically in terms of perceived violations with respect to the individual in question) (Colquitt et al., 2001; Tyler and Blader, 2003).

The theory of ego depletion has recently been posed by social psychologists to understand poor decision making. It sees the availability of individual willpower or self-control as varying over time as a function of factors such as sleep deprivation, low blood glucose, or resource scarcity (Baumeister and Tierney, 2011; Baumeister et al., 2000; Gino et al., 2011; Mani et al., 2013).

Kahneman and Tversky's (1979) Prospect Theory addresses decision making under uncertainty, focusing attention on the "bounded rationality" of actors. This theory appears particularly relevant due to the parallel ideas that, on the one hand, fear of loss (loss aversion) tends to be a much stronger motivator of behavior than does the potential for gain, and on the other hand that individuals tend toward risk aversion when confronted with potential gains but bias toward risk seeking when confronted with avoiding potential losses. Applied to a research setting, this theory would imply that, other things being equal, researchers facing a potential loss of position or resources would be more inclined to take risks—including research misconduct or DRPs—than those seeking to gain status or resources.[1]

Experimental work in the social and behavioral sciences has shed light on how these theoretical perspectives can be applied to specific problems such as cheating by students that could carry implications for research practices. Ariely and his colleagues, in a series of experiments, have found that the extent to which human beings are willing to cheat and engage in dishonest behavior "depends on the structure of our daily environment" (Ariely, 2012). A key finding is that maintaining a self-image of honesty is important to people, but many are able to engage in very low levels of cheating and simultaneously adjust their explanations to retain their own self-regard (Mazar et al., 2008).

A recent compilation of decades of research on cheating by students, for example, focused on five elements that combine to contribute to an environment that is conducive to cheating: (1) a strong emphasis on performance, (2) very high stakes, (3) extrinsic motivation, (4) a low expectation of success, and (5) a peer culture that accepts or endorses corner cutting or cheating (Lang, 2013). Among the top reasons that students use to rationalize cheating is when the teacher/assessment system is perceived to be unfair and/or there is perceived to be little chance of success (Brent and Atkisson, 2011). This tracks to the literature on

[1] This is not meant to imply that researchers facing a potential loss would choose risky research topics. Indeed, such a researcher might exhibit risky behavior in the form of misconduct while working on a topic currently popular in his or her field.

organizational justice showing that when humans perceive a workplace as arbitrary and unfair, they find it more justifiable to (and are more likely to) cheat by stealing or by calling in sick than when the workplace is perceived as being fair (Folger and Cropanzano, 1998).

### Implications for Understanding and Preventing Research Misconduct and Detrimental Research Practices

The foregoing sketches out an evolution of thinking over time—supported by empirical work—from a focus on deviance as stemming from a bad actor's rational set of choices to a more nuanced understanding of the multifactorial influences on human decision making. Humans are influenced by a wide range of cognitive biases and errors that infect and disrupt rational thought—even when we think we are being rational. No single theory is likely to be adequate to completely explain the full spectrum of behavior encompassed by research misconduct and detrimental research practices, but to assume that researchers are not subject to the same kinds of influences and defects in their decision making that afflict humans more generally would also be a mistake.

Some preliminary, limited research has attempted to bring some of this theoretical richness to bear directly on the questions of research misconduct and detrimental research practices, but the existing research has not been well positioned to provide compelling tests of the hypotheses suggested by these perspectives (Antes et al., 2007; DuBois et al., 2016; Martinson et al., 2006, 2010; Medeiros et al., 2014; Mumford et al., 2007, 2008). An analysis of research misconduct case files showed that a variety of causes and rationalizations could be identified, including personal and professional stressors, organizational climate, and personality factors (Davis et al., 2007). Generating more precise insights and more adequate tests of the theoretical frameworks useful for understanding research misconduct and detrimental research practices would require far more detailed longitudinal, perhaps experimental (and perhaps social-network–based, in some cases) data than have been amassed and examined in the study of research integrity.

While questions of the prevention of research misconduct are implicit in most, if not all, of the theoretical perspectives discussed above, some who have studied the topic have been more explicit in differentiating perspectives on prevention. In 2005, Douglas Adams and Kenneth Pimple offered a criminological perspective on the topic of prevention of research misconduct, arguing that any instance of misconduct can be described as having two essential elements—a propensity on the part of the individual to engage in a deviant behavior, and the opportunity to do so (Adams and Pimple, 2005). This is a somewhat different take on the joint person/environment explanatory framework seen in other theories, and these authors argue that anticipated difficulty in altering the propensity of individuals to deviate from norms suggests that opportunities for misbehavior

should be reduced. They do not address the topic seen in many other theories of trying to address motivations for deviant behavior. This line of work is largely consistent with situational prevention approaches in criminology that seek to increase the risks and costs of specific categories of crime while reducing rewards through manipulation of the environment and other techniques (Clarke, 1995).

In contrast to a narrow focus on misconduct, Nylenna and Simonsen have offered an epidemiologic perspective that draws attention to the entire distribution of behavior composed of research misconduct and detrimental research practices (Nylenna and Simonsen, 2006). These writers start from Geoffrey Rose's now-classic population health perspective, which argued that when a disease risk factor is widespread in a population (e.g., hypertension), attempting to bring the entire risk distribution down should be the objective: reducing the overall risk distribution only slightly may be a more effective approach than simply trying to eliminate the risk only in the most "high risk" individuals (Rose, 1985). That is, Rose argued that reducing blood pressure population-wide by only a few mm Hg is a more effective way to reduce heart disease than merely bringing the blood pressure of a smaller group of hypertensives below the 140/90 threshold. Nylenna and Simonsen's insight was that, like disease risks in a population, there is a range of DRPs in science beyond the most extreme examples of misconduct that meet the federal definition, and that focusing on the broader range of undesirable, research-related behavior might be more beneficial than a single-minded focus on fabrication, falsification, and plagiarism.

Like Nylenna and Simonsen, Weed brings an epidemiologic perspective to the prevention of research misconduct (Weed, 1998). Weed distinguishes three types of prevention, analogous with concepts of prevention in medicine: primary prevention, secondary prevention, and tertiary prevention. He sees primary prevention as "identifying and removing causes of events and as identifying factors whose presence (rather than absence) actively reduces the occurrence of those events" (Weed, 1998). He discusses secondary prevention as early detection to increase opportunities for discovering instances of misconduct and "treatment" through "procedures for investigating cases as well as the sanctions delivered to those responsible for the misconduct" (Weed, 1998). Auditing and increased monitoring of junior researchers are cited as examples of secondary prevention. In terms of tertiary prevention, Weed suggests that it "can also be applied to scientific misconduct, inasmuch as those who commit such misconduct may require rehabilitation before they return to scientific practice" (Weed, 1998).

Weed notes the difficulties in knowing anything about scientific misconduct with any level of certainty, owing to multiple factors, including the hiddenness of the behaviors in question, but also the lack of existing data and general absence of resources devoted to their study. In particular, he notes:

> Indeed, in the foregoing analysis, a host of such questions have emerged. Answers will be difficult to obtain, especially if precise scientific methodologies are to be employed. But then, we are scientists, and solving difficult empirical

> problems is what we do best. Perhaps the essential question is less methodological than motivational: Are we as scientists willing to study our conduct as scientists? If so, then one day we may discover why we suffer from an important and sometimes disabling professional affliction and what works to prevent it.
>
> I am not suggesting, however, that we should postpone interventions until we fully understand the etiology, including the underlying biological, behavioral, and social mechanisms involved in the range of activities we call scientific misconduct. (Weed, 1998)

Reason's work on high-reliability systems offers a framework for considering certain behaviors as possibly amenable to being addressed through quality improvement and quality assurance mechanisms, and that doing so may be a more effective way of reducing undesirable behavior than the historical focus on criminality (Reason, 2000). Such a perspective may be particularly salient in considering the potential role of sloppy research practices in contributing to the reproducibility problem, which is discussed in Chapter 5 and Chapter 7. Reason distinguishes a "person approach" to error from a "system approach" (Reason, 2000). He notes that "high reliability" organizations (including air traffic control centers, nuclear power plants, and nuclear aircraft carriers) are characterized by a focus on error management at the systems level more than the individual level. Reason's logic has also been adopted in widespread attempts to reduce medical errors and a focus on moving medicine from a "blame and shame" cultural perspective to one of a "reporting and feedback" cultural perspective, with public reporting of individual and organizational performance being crucial (Leape, 2010). There is potential value in bringing such a perspective to bear, particularly in the promotion of research best practices and other quality improvement efforts recently being considered to improve the reliability and reproducibility of research findings, as discussed in more detail in Chapter 7.

## CURRENT FUNDING AND ORGANIZATIONAL TRENDS AND THEIR NEGATIVE IMPACTS ON RESEARCH ENVIRONMENTS

Research environments at institutions and laboratories that produce outstanding work have long been characterized by significant competitiveness and pressure to perform. However, patterns of funding and organization that have emerged over the past few decades in the United States have created environments increasingly characterized by elements identified above that are associated with cheating, such as very high stakes, a very low expectation of success, and peer cultures that accept corner cutting. These conditions are best documented in the single largest component of the research enterprise in the United States—biomedical research—but aspects of these problems are appearing in other disciplines (Alberts et al., 2014; Casadevall and Fang, 2012; Stephan, 2012b; Teitelbaum, 2008).

Data strongly suggest that some fields have been producing more highly trained students with specialized research training than can be absorbed in research. According to the report of a National Institutes of Health (NIH) working group that examined the biomedical research workforce, the production of PhDs in the biomedical sciences has closely tracked the NIH budget (NIH, 2012b). Thus, the number of basic biomedical PhDs began to increase substantially in 2004, just as the doubling of the NIH budget was ending, which reflects the 5- to 7-year graduate student cycle. Over the past two decades, the number of basic biomedical graduate students has doubled (NIH, 2012b). Yet the percentage of PhDs who move into tenure-track positions has dropped from 34 percent in 1993 to 26 percent today, and the percentage of graduates in the biomedical sciences who say that they are employed in occupations closely related to their PhD field dropped from 70 percent in 1997 to 59 percent in 2008 (NIH, 2012b).

Across all of science and engineering, annual production of PhDs in the United States is roughly 10 times the number of open faculty positions (Schillebeeckx et al., 2013). Figure 6-1 shows the ratio of younger researchers who hold faculty positions to those holding postdoctoral fellowships and other temporary positions and how that ratio has declined over time. As discussed further below, the purpose of examining and highlighting labor market trends for PhD scientists and changes is not to imply that all or most science and engineering PhD recipients should ultimately be employed in academic research. Yet the scope of the discrepancy must lead us to ask: For what careers are all of these PhDs being trained?

In addition to the low probability of success, achieving a tenure-track position is a lengthy process. While the time to degree and age at degree have remained stable over the past 15 years, the overall length of training in the biomedical sciences, including graduate school and postdoctoral appointments, is longer than for comparable disciplines, and those who go on to tenure-track positions do so at an older age than PhDs in other disciplines. The average age at which PhD biomedical investigators get their first R01 grant from NIH is 42, an age at which investigators in other disciplines, not to mention nonacademic professionals, typically are already well established in their careers. The overall age profile of principal investigators has also risen.

This is not to say that all of those who have completed the education and training required of biomedical research principal investigators necessarily want to or have to go into research. Some medical schools have made it a high priority to train their PhD candidates for a variety of careers over the past decade. And students want other kinds of careers. Some institutions do a commendable job of communicating career information to students. Many trained researchers can and do find rewarding and successful careers in industry, government, and nonprofit organizations.

However, PhD scientists have received specialized training for this work, and academic research has traditionally employed a significant percentage of bio-

medical PhD recipients. If the number of PhDs rises while the number of tenure-track positions goes down, this implies that there will be heightened competition for those positions and that a growing fraction of PhD scientists will need to find employment in other sectors. The education and training requirements for some alternative careers (such as a professional science master's) require fewer years of education and training than a tenure-track position in biomedical research. In aggregate terms, the end result is a system that relies more heavily on postdocs, graduate students, and other nonfaculty researchers relative to faculty than it did in the past.

Surveys have been taken to better understand the causes and consequences of apparent workforce imbalances in academic research and the challenges facing early-career researchers, with particular attention focused on the postdoctoral experience (Sauermann and Roach, 2012). For example, are imbalances primarily caused by supply-side or demand-side conditions? A 2014 report proposed both supply-side interventions, such as providing better education to graduate students and postdocs about career possibilities, and demand-side measures, such as limiting the length of postdoctoral service and raising minimum salaries (NAS-NAE-IOM, 2014). The evidence from surveys indicates that this is a complex issue. Graduate students and postdocs may have an accurate understanding of the abstract probability of attaining a tenured research position in their fields but may overestimate their own chances (Sauermann and Roach, 2012).

Even for the minority of researchers in biomedical fields who are eventually able to secure tenure-track faculty positions, the research environment continues to be characterized by hypercompetition. The success rate for NIH grant applica-

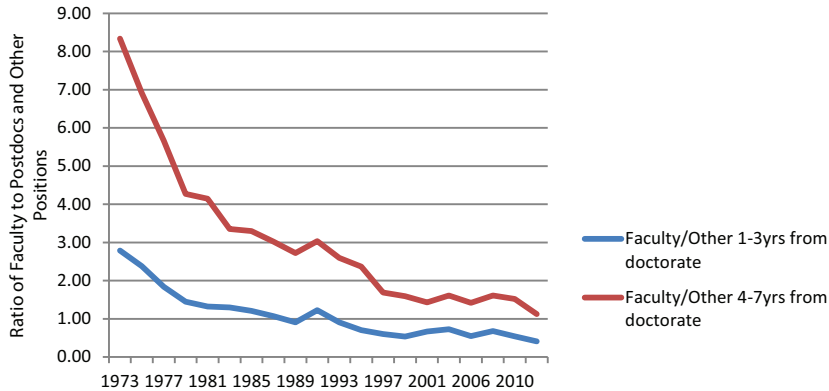


FIGURE 6-1 The ratio of faculty members to postdocs and other science, engineering, and health doctorates employed in academia has declined over time.
SOURCE: Data taken from NSB, 2016.

tions has fallen from 32 percent in 2001 to 18 percent in 2015. The average size of grants has risen in current dollars but has declined in real terms between 1999 and 2015 (NIH, 2015). The success rate at the National Science Foundation has experienced a more modest decline, from 27 percent in 2001 to 20 percent in 2014, with the size of the average award not showing a significant trend (in real dollars) over that period (NSB, 2014a, 2016). As a consequence, faculty members have to spend more time writing grants and less time doing research.

While success rates have been declining, external funding has become an important source of coverage for faculty salaries. To be sure, charging faculty salaries to grants is a proper and legitimate expense for federally funded work. And reviewers might question a principal investigator's commitment to a project if his or her salary allocation is too low. At the same time, the emergence of a situation at some institutions and departments where the salaries of a substantial fraction of the research workforce are heavily or even entirely dependent on grant funding is a relatively recent phenomenon.

A survey undertaken by the Association of American Medical Colleges showed just under half of the salary support for non-MD faculty at responding institutions came from sponsored programs in fiscal 2009 (Goodwin et al., 2011). Aggregate data obscure the fact that conditions may vary greatly for individuals within an institution, with some faculty completely dependent on grant funding. They also mask differences between institutions. For example, 2013 Association of American Medical Colleges data from 72 academic medical centers showed that while most institutions fell in a range where 40 to 60 percent of non-MD faculty salaries were covered by sponsored programs, a minority of institutions (fewer than 10) relied on sponsored programs for up to 70 percent of non-MD faculty salaries (Levine et al., 2015). Anecdotal evidence reveals that reliance on sponsored programs for salary coverage is a considerable source of stress and anxiety at some institutions, particularly for younger faculty (Hellweg, 2015).

Some experts have pointed to the reliance on "soft money" faculty positions supported by grant funding as an important indicator that incentives are not aligned to support high-quality research (Alberts et al., 2014; Stephan, 2012b; Teitelbaum, 2008). The potential drawbacks of such salary arrangements have also been questioned by the NIH leadership (Collins, 2010). The thinking goes that researchers who are dependent on grant funding to support their salaries will be inclined to propose safer research that is more likely to receive funding but less likely to result in significant advances (Stephan, 2012b). This concern was actually raised in a 1960 report of the President's Science Advisory Committee, which warned of "the need for avoiding situations in which a professor becomes partly or wholly responsible for raising his own salary" (PSAC, 1960).

A study that examined differences between researchers funded by the Howard Hughes Medical Institute, "which tolerates early failure, rewards long-term success, and gives its appointees great freedom to experiment," and grantees of NIH,

"who are subject to short review cycles, predefined deliverables, and renewal policies unforgiving of failure," appears to bear this out (Azoulay et al., 2011).

As noted above, an analysis of actual research misconduct cases implicates factors related to a hypercompetitive research environment in many of the cases (Davis et al., 2007). In another study conducted through a series of focus groups with early- and mid-career scientists, Anderson et al. (2007b) found that competition in research "can skew this system in unanticipated and perverse ways, with negative consequences for science as well as for the lives and careers of scientists." Several of the interviewees pointed to a dramatic increase in the competitiveness of research in recent years. As one said, "It's so much more competitive than it used to be. When we were first starting out, it was more collegial. You gave reagents away freely. Now there's more at stake. There's patents at stake. There is getting yourself funded."

One of the negative consequences of competition cited by scientists is an inducement to engage in careless or detrimental research practices. While none of the interviewees said that they had committed or witnessed misconduct in science, they cited the temptation to behave irresponsibly. As one said, "There is a lot of pressure for people to come out with things in a very short time-frame. The likelihood that corners are cut, is real." These incentives can operate internationally, as some institutions and governments provide large bonuses to researchers who are able to publish work in certain prestigious international journals. Also, empirical findings have shown a strong positive relationship between the level of competition perceived in an academic department and the likelihood of misconduct being observed by departmental members (Louis et al., 1995).

Appendix D discusses a specific case in which Elizabeth Goodwin, a faculty member at the University of Wisconsin who oversaw a number of graduate students and postdocs, was found to have falsified information contained in a federal grant application she submitted in 2005 (ORI, 2010). No evidence has emerged indicating that Goodwin committed research misconduct in other applications or in her publications. Philip Anderson, a faculty member in Goodwin's former department, took one of the graduate students into his lab to finish work on her PhD dissertation. Anderson provided his perspective in an interview several years later: "I've thought about Betsy a lot through this process…. What she did, I believe, happened because of the extreme pressure we're all under to find funding" (Allen, 2008). Certainly, this is one opinion, and it is not cited to rationalize or justify misconduct. However, as discussed above, there is empirical and theoretical grounding for concerns about the potentially detrimental effects of competition for resources on the behavior of scientists.

Recalling the discussion from Chapter 1 that described the research enterprise as a complex system, the accumulating evidence seems to suggest that some aspects of this system are not functioning well, at least in some disciplines, with implications for the ability of the system's stakeholders to foster research integrity. If producing the next generation of highly trained and educated scientists

is an important function of the system, then there is reason to question whether current funding policies and the incentives that they create for institutions and individuals are resulting in the right quantity or quality of that output. The structural challenges facing U.S. biomedical research and the resulting perverse incentives and unintended negative consequences have been described by leading scientists (Alberts et al., 2014; Casadevall and Fang, 2012). These scientists have linked structural challenges to incentives affecting whether researchers commit misconduct or DRPs and assert that they will not be solved simply through increased funding. While developing approaches to remedy these structural challenges is beyond the scope of this report, the linkages between structural features of the research enterprise such as funding mechanisms, research environments, incentives, and behavior need to be better understood.

## THE VALUE AND IMPORTANCE OF RESEARCH ON RESEARCH INTEGRITY

The discussion in this chapter illustrates that the causes of research misconduct and detrimental research practices are complex. The current state of knowledge, while incomplete, supports several propositions:

1. Rather than focusing exclusively on fabrication, falsification, and plagiarism, the breadth of research misconduct and detrimental research practices should be taken into account and addressed.
2. Targeting of efforts needs to go beyond just bad actors and should attend to the salient features of both local organizational environments and settings, as well as the structural arrangements of research and the incentive structures with which various actors in the research enterprise are confronted.
3. Additional theoretically grounded research is warranted to more completely inform all such efforts. Important areas of focus include why researchers commit misconduct and DRPs—both theoretical and empirical work—as well as the strategies and interventions that could prevent or reduce the incidence of these actions.

These propositions underlie the committee's Finding C, discussed in Chapter 11, which addresses the need for more knowledge to develop evidence-based approaches to research misconduct and detrimental research practices.

# 7

# Addressing Research Misconduct and Detrimental Research Practices: Current Knowledge and Issues

*Public policy on research misconduct, which has developed contentiously in the United States and a few other countries over the past thirty years, remains largely untested as to whether it yields clearly specific outcomes; alternative policies that might reach those outcomes remain unexamined. Each widely publicized case of research misconduct creates a new scandal, leading to questions about whether current regulation is effective or just, and whether it supports the progress of science.*

—Barbara Redman (2013)

***Synopsis:*** *Research misconduct and detrimental research practices are addressed in several ways. Addressing misconduct and detrimental research practices through the implementation of standards and best practices, such as effective mentoring at the lab level, requirements for data and code sharing at the disciplinary level, and implementation of greater transparency in reporting results, can strengthen the self-correcting nature of science. Efforts to prevent them through education are described in Chapter 10. In the United States, uncovering, establishing, and responding to misconduct in publicly funded research mainly takes place within the context of the federal research misconduct policy. The current policy framework assigns specific responsibilities to institutions and to sponsoring agencies. While the current framework has achieved stability and effectiveness in ensuring that misconduct allegations involving federally funded work are investigated, there are gaps and inconsistencies. Other countries have different policy frameworks for addressing misconduct, which has implications for the United States due to the growing number of international research collaborations. Addressing detrimental research practices may involve even greater challenges than does addressing misconduct.*

## ADDRESSING RESEARCH MISCONDUCT

Chapter 4 provides a broad overview of how the U.S. policy framework for addressing research misconduct has evolved and describes some basic elements of that framework, most notably the federal definition of research misconduct.

Relevant international developments and policies are also covered. To assess the strengths and weaknesses of current approaches, it is necessary to explore how the policy framework operates in practice.

Several decades ago, a fairly widespread viewpoint among scientists was that federal policies to deal with research misconduct were not necessary, since misconduct was extremely rare and the self-correcting nature of science would ensure that any misconduct would be quickly discovered (Gunsalus, 1997). One basis for this viewpoint was the social cohesion of research fields and subfields at a time when researchers were much likelier to know each other than is the case today. In an environment where personal relationships and reputations play an important role in professional success and advancement, senior researchers have strong incentives to be effective mentors, since successful students would enhance their reputations. Likewise, if one's current or former student is caught plagiarizing or fabricating data, the mentor or supervisor's reputation will suffer.

While subfield communities still play a very important role in research, and misconduct certainly causes supervisors and collaborators to suffer embarrassment, it is unrealistic to think that these social forces are a sufficient deterrent to actual misconduct. As described in Chapter 3, the conditions in which less formal approaches to fostering integrity or to uncovering and addressing misconduct might have been expected to effectively protect the research record and the health of the research enterprise, to the extent that they ever existed, are certainly long gone.

Education in the responsible conduct of research (RCR) is another mechanism for addressing research misconduct that has been widely advocated for the prevention of misconduct and detrimental practices. Chapter 10 features an extensive discussion of RCR education and what is known and not known about its effectiveness and benefits. The federal mandates related to RCR education reflect the logic that a significant percentage of research misconduct might be committed due to a lack of understanding, and that addressing this could prevent some research misconduct. Certainly one can appreciate funding agency frustration with cases in which an early career respondent might claim that he or she was never taught that behavior such as copying large blocks of text from other work is wrong, and where the institution counters that it did train the respondent not to plagiarize. RCR education mandates might at least prevent or significantly reduce claims of ignorance as to the basic values and practices of science. In addition, as discussed in Chapter 10, while most experts and practitioners believe that RCR education is necessary and worthwhile, perhaps particularly in discouraging detrimental research practices, the evidence of its effectiveness is limited. This issue is discussed in more detail in Chapter 10 and in Appendix C.

It is possible that RCR education has prevented some number of acts of research misconduct. However, the experience of the past several decades shows that it may be insufficient to rely on classroom or online education as the primary tool to address research misconduct.

As described in *Responsible Science*, before the mid-1980s, allegations of research misconduct were investigated and addressed by institutions. Institutions would employ a variety of procedures, sometimes confidential, with no requirement that the institution notify the sponsoring agency of the investigation or the results (NAS-NAE-IOM, 1992). Federal policies instituted since that time have required research institutions to report to the sponsoring agencies when initial inquiries yield enough evidence to justify a full investigation. According to the current federal research misconduct policy

> Agencies and research institutions are partners who share responsibility for the research process. Federal agencies have ultimate oversight authority for Federally funded research, but research institutions bear primary responsibility for prevention and detection of research misconduct and for the inquiry, investigation, and adjudication of research misconduct alleged to have occurred in association with their own institution. (OSTP, 2000)

For misconduct allegations covering research sponsored by federal agencies such as the National Science Foundation (NSF) and the National Institutes of Health (NIH), institutions are responsible for notifying the cognizant offices—NSF's Office of the Inspector General (NSF-OIG) for NSF-funded research, and the Department of Health and Human Services' (HHS's) Office of Research Integrity (ORI) for research funded by NIH and other Public Health Service (PHS) entities—when investigations are launched and when they conclude. Differences between NSF-OIG and ORI in how cases are addressed, as well as issues that arise for misconduct alleged in research sponsored by other agencies, will be explored in more detail below.

As described in Chapter 5, much remains unknown about the incidence of research misconduct and trends. For example, how many cases of misconduct go unreported and/or are not investigated is unknown. In addition, detailed information is lacking about the circumstances of many cases where misconduct has been found. For example, ORI posts summaries of the cases where "administrative actions were imposed due to findings of research misconduct" (ORI, 2015), and NSF-OIG provides a searchable database of closeout memos from all of its investigations, including research misconduct investigations (NSF-OIG, 2015). These case descriptions may not include information about how the misconduct was uncovered and other details that could be useful to institutions and other stakeholders seeking to improve approaches to preventing misconduct and to uncovering the misconduct that does occur. As a result, cases that have achieved enough notoriety to attract media reports tend to be the primary source of information about how research misconduct is uncovered and addressed.

Research misconduct cases regularly emerge in the current environment, including investigators who have fabricated data underlying tens or even hundreds of publications over the course of lengthy careers. Dealing with allegations and

correcting the research record are significant activities for institutions, sponsors, journals, and other stakeholders.

## Uncovering Research Misconduct

Examining how allegations arise and are dealt with provides a useful window into the system for addressing misconduct, making it clear that science's broad tendency toward self-correction or other mechanisms such as traditional prepublication peer review cannot be relied on as the primary mechanisms for uncovering misconduct. Such examination also provides insights on some of the system's weaknesses and some possible clues as to how the system might be strengthened.

The discovery of research misconduct often depends on good-faith whistleblowers who observe the wrongdoing and come forward to report it. In a study by Stroebe et al. (2012), an examination of 40 "notorious" cases of research misconduct from 1974 through 2012, defined as those cases that were prominent enough to receive media attention and where the mode of discovery could be ascertained from media or other reports, found that about half were uncovered due to whistleblowers. Only a few were uncovered through a failure to replicate or in the process of peer review. Several of the cases included in Appendix D are included in the sample.

Another analysis examined cases of research misconduct as well as cases of other misconduct—such as failure to follow rules governing human subjects or laboratory animal protection—and medical practice misconduct (e.g., undertaking unnecessary procedures) (DuBois et al., 2013b). Of the research misconduct cases analyzed, 28 percent involved a "failed attempt at reporting research misconduct (i.e., the wrongdoing continued for some time following an initial report)." A large percentage of research misconduct whistleblowers worked within the wrongdoer's institution, including 23 percent who were subordinates.

While the actions of whistleblowers play a central role in uncovering research misconduct, particularly in cases of data fabrication, they are not the only way that misconduct is discovered and reported. For example, several technological methods of detecting misconduct have emerged over the past decade. A well-known example is software that detects and flags text overlap that can then be checked for possible plagiarism. It is important to use this tool in a sophisticated way, since standard equations and citations may set off the software. According to one report, scientific publishers that began screening submissions on a trial basis in 2010 found from 6 to 23 percent of articles had to be rejected due to unacceptable levels of text duplicated from other articles, depending on the journal (Butler, 2010). Software has also been developed to detect the inappropriate manipulation of image data, a form of data falsification that has emerged regularly in the life sciences (Rossner, 2006).

In addition, Uri Simonsohn of the University of Pennsylvania has developed a methodology that enables the detection of fabricated or falsified data

through the analysis of datasets (Shea, 2012; Simonsohn, 2013). The methodology uses statistical analysis to determine the probability that a given large dataset was generated by an experiment as opposed to being fabricated. Using this methodology, Simonsohn uncovered data fabrication by University of Michigan psychology professor Lawrence Sanna and by Erasmus University (Netherlands) psychology professor Dirk Smeesters. Methods for analyzing clinical data to detect fabrication predated Simonsohn's efforts, having appeared in the late 1990s (Buyse et al., 1999; Lock et al., 2001). Statcheck is a new tool, developed at Tilburg University in The Netherlands, that can check the reported statistical results in articles for consistency (Epskamp and Nuijten, 2016).

Technological approaches to detecting research misconduct are not foolproof. For example, inventive fabricators might be able to devise ways of defeating statistical analysis of datasets. Still, the existence and use of these tools is encouraging. Improving on them and building new approaches to detecting misconduct will rely heavily on improved transparency throughout the research process, particularly in the availability of data and code. For example, the effectiveness of Simonsohn's method depends on access to data, and on the fact that fabricated data will differ from data generated by an experiment in discernible ways. The importance of increasing transparency is a key theme of this chapter and underlies several of the committee's recommendations intended to prevent or reduce misconduct and detrimental research practices (DRPs) and to more effectively detect the misconduct that does occur.

As discussed above and in Chapter 5, the failure to replicate work has historically not been a primary means for uncovering misconduct, for several reasons. First, while there are incentives for extending and building on previous work, the incentives to replicate work that has already been reported are weak. In addition, the standards of some fields for sharing data and methods may not currently be robust enough to ensure that all the information necessary to replicate or validate a study is provided. In the biosciences, in particular, it may be difficult to account for a variety of nuances that may be important in result replication.

However, in recent years there have been several cases where doubts or suspicions about groundbreaking or otherwise newsworthy results have appeared almost immediately, leading to the findings either quickly falling apart or prompting more thorough investigations. For example, two papers by Haruko Obokata of Japan's RIKEN research institute and an international group of coauthors on reprogramming mature stem cells into embryonic stem cells by using an acid bath were published by *Nature* in 2014 (Obokata, 2014a,b). Within a few weeks, outside researchers who were unsuccessful in replicating or extending the work were questioning the results (Cyranoski, 2014b). A RIKEN investigation found that Obokata had intentionally falsified data (Ishii et al., 2014) (see Appendix D). In a second example, a paper published in *Science* in late 2014 purported to show that canvassers could be highly successful in changing the minds of voters opposed to same-sex marriage, in many cases with a single conversation (LaCour

and Green, 2014). The paper was subsequently retracted after replication efforts failed and one of the authors, a graduate student in political science at the University of California, Los Angeles, admitted to destroying the raw data, leading to an investigation (McNutt, 2015).

One analysis of retractions indicates that the frequency of retractions is positively associated with a journal's impact factor—that is, that more prestigious journals have to retract articles at a higher rate (Fang et al., 2012). This may be partly due to the perceived risk-reward balance for potential fabricators (the higher rewards that come from publishing in a high-prestige journal lead to stronger incentives to cheat). In addition, articles in high-prestige journals will generally receive greater attention and scrutiny, implying that misconduct is more likely to be discovered. Perhaps, in some cases, pressure to expedite publication means that corners are cut in the review and publication processes. These journals may also be more sensitive to the need for timely retractions and have greater resources to investigate issues. The cases from recent years are somewhat encouraging in illustrating that self-correction in science can work where the community has sufficient information and where outside researchers have strong incentives to replicate and extend the work.

### Uninvestigated Misconduct

The extent of research misconduct that is never uncovered, reported, or investigated is unknown by definition. Chapter 5 discusses the existing evidence on the incidence of research misconduct. Surveys of researchers on their own behavior and on the behavior of colleagues whom they have observed or heard about generate much higher estimates of the incidence of research misconduct than is reflected in the findings of research misconduct investigations reported by NSF-OIG and ORI. Over time, surveys have become more sophisticated in addressing issues that would tend to inflate the reported incidence of misconduct, such as possible multiple counting of the same incident by different respondents. According to one assessment, the majority of misconduct cases are not reported (Titus et al., 2008). The number or percentage of research misconduct cases that are not investigated cannot be pinpointed; nevertheless, it is important to try to understand as much about these cases as possible.

In addition to not knowing the true incidence of research misconduct, the circumstances and outcomes of research misconduct cases that may be reported or detected but may not be officially investigated remain largely unknown. Yet some useful information does exist. For example, there are anecdotal accounts by journal editors of what they have done when they, their peer reviewers, or outside whistleblowers have raised concerns and suspicions about submitted work (White, 2005). These accounts illustrate what can happen at different points in the process to forestall an investigation, other than the journal receiving a clarification or additional information that allays the suspicion. They also illustrate

some of the reasons why the journal peer review and editorial processes are not as effective in uncovering misconduct as might be expected or hoped for.

For example, the guidelines of the Committee on Publication Ethics specify that journal editors should "inform institutions if they suspect misconduct by their researchers, and provide evidence to support these concerns" (Wager and Kleinert, 2012). While journal editors are not equipped to actually perform investigations themselves (Wager, 2015a), journals are advised to go to the authors first and to contact the institution if the response is inadequate. In cases where suspicions have been raised, editors may not believe that they have sufficient evidence to go to the institution, and this determination may depend on the experience and attitudes of the editor. One editor reported having been hesitant to raise suspicions with institutions early in his career, unless he had compelling evidence of misconduct, but had become less hesitant over time (Smith, 2006).

In cases where the journal editor goes to the institution but the institution does not reply, the editor would have no way of knowing whether the institution had undertaken a preliminary inquiry, proceeded to a full investigation, or not taken any action. In these cases, journal editors are advised to be persistent and to contact the funder or national research integrity office if the institutional response is inadequate. In some countries, institutions do not have a formal responsibility to investigate research misconduct or report it to sponsors, and there may be no national research integrity organization. The journal may have little or no ability in these cases to put pressure on institutions to respond, or even to prevent the author from submitting the article to another journal with less rigorous editorial practices after rejecting it. Journals might notify institutions that do not respond to credible concerns and allegations that future submissions from researchers affiliated with the institution would not be considered for publication until the issue was addressed. Obviously, more prestigious journals are likely to find greater success with this approach than less prestigious ones. Success may also depend on the institutional official who has been notified. Journals might also refrain from such an approach because it punishes the innocent along with the guilty.

Journals also have the responsibility to respond to institutional requests to retract fabricated or falsified work, and they sometimes fail in this responsibility. Retractions and related issues are discussed in Chapter 5 (Wager, 2015b).

A 2015 report examined 57 published clinical trials undertaken over the period 1998–2013 in which Food and Drug Administration (FDA) inspections of clinical trial sites had found significant evidence of one or more problems, such as protocol violations, inadequate record keeping, and failure to protect patient safety (Seife, 2015; Steinbrook and Redberg, 2015). Significant evidence for falsification or submission of false information was found in 39 percent of the trials. However, of the 78 publications resulting from the 57 clinical trials, only 3 mentioned the problems that had been discovered in the FDA inspections. It is not clear from the inspection documents or the publications how the problems were communicated to institutions and whether inquiries or investigations of research

misconduct or violations of human subjects protection regulations were ever performed. The cases where evidence of falsification was found potentially represent examples of research misconduct being uncovered "in the act," so to speak, prior to publication, and not being investigated, with the results published as if nothing had happened. In 2012 the FDA published new regulations strengthening its ability to disqualify clinical investigators who falsify data or commit other violations (HHS, 2012). The existence of these regulations does not necessarily ensure that the findings of FDA investigations that appear to justify research misconduct inquiries or investigations on the part of institutions are followed up. To be sure, the legal authorities and implementing regulations that govern how FDA exercises its responsibilities have evolved separately from the federal research misconduct policy and related regulations, so it is not surprising that they might be out of synch in some areas. As indicated in this discussion, not much is known about how these policy frameworks interact in practice and what sorts of changes or adjustments might be needed.

### Blogs, Websites, and Community Postpublication Review

Over the past few years, several blogs and websites have emerged that focus on research misconduct and related issues. The best known of these efforts is the blog *Retraction Watch*, which was launched by two science journalists in 2010 (Oransky and Marcus, 2010). The blog's authors expressed several goals in starting it, such as gaining a better understanding of the scientific process, serving as an informal repository and notification site for retractions, providing information to journalists seeking to uncover research misconduct, and evaluating the performance of journals. *Retraction Watch* has gained a wide readership within the research enterprise.

The effectiveness and impact of *Retraction Watch* have not been formally evaluated, but it is plausible to argue that the blog has advanced the goals that the authors set out for it. For example, the specific mechanisms by which retractions are communicated and retraction notices are maintained by journals are not standardized. Since retractions have not traditionally been widely publicized, prior to the emergence of *Retraction Watch* it was possible that an individual retraction might not be noticed immediately by other researchers in the field or by other journals that have published work by the authors in question. This could delay examination of other work by those authors and correction of the literature. In those cases where authors have fabricated or falsified data in multiple papers, having a report of a retraction appear in *Retraction Watch* can accelerate this process of examining the researchers' broader body of work. It might be possible to look at cases that emerged before and after the advent of *Retraction Watch* in order to establish or quantify possible effects. Other issues related to retractions are discussed further below and in Chapter 5.

Besides drawing greater attention to issues and barriers in publication prac-

tices that may delay retractions or prevent a clear explanation of the cause, *Retraction Watch* has highlighted some of the information deficits around research misconduct and detrimental research practices, such as a lack of data on some issues that cover all disciplines. For example, much of the recent literature that examines retractions relies on searches of retraction notices in PubMed, which focuses on the biomedical literature and does not comprehensively cover the physical sciences (e.g., Fang et al., 2012).

Several web-based initiatives have aimed at facilitating the discussion of suspicious publications and uncovering research misconduct. For example, the website Science-Fraud.org was operated by Paul Brookes, a medical professor at the University of Rochester, during 2012 (Couzin-Frankel, 2013). The site provided a forum for reporting and discussing suspicious images in published work. Brookes and the contributors to the site operated anonymously, and Brookes claims that information provided on the site led to 16 retractions and 47 corrections (Pain, 2014). However, Brookes shut the site down in early 2013 after his identity was revealed in an e-mail sent to his university and many of the researchers whose work was questioned on Science-Fraud.org. The strident tone of the website, which went beyond raising questions about published work to accusing researchers of misconduct, opened Brookes to threats of legal action.

Another example is the website PubPeer, which provides a forum for commenters to critique published work and is moderated anonymously. Content made available on PubPeer has also led to corrections and retractions. In 2014, Fazlul Sarkar, a cancer researcher at Wayne State University, sued several PubPeer commenters on his papers, claiming that their posts constituted defamation and caused him to lose a job offer. Sarkar sought identifying information on the commenters from PubPeer via subpoena (Servick, 2015). It was later revealed that a tipster who had raised concerns and issues regarding numerous journal articles with editors over the years and goes by the pseudonym "Clare Francis" was one of the PubPeer commenters on Sarkar's work (Oransky, 2015). While Francis's communications have sometimes led to retractions or corrections, journal editors have also asserted that some of the tips did not actually uncover mistakes or wrongdoing and that investigating them wasted time (Grens, 2013b).

The phenomenon of websites such as PubPeer and whistleblowers such as Clare Francis raise questions about the role of anonymous whistleblowers and about how the community, and journals in particular, should treat such accusations and concerns. The topic of knowingly false allegations is discussed below. Journal editors need to exercise judgment in evaluating the credibility of expressions of concern and accusations they receive, and anonymity deprives them of important information in making an evaluation. However, the desire for anonymity on the part of whistleblowers is also understandable, particularly in cases where exposure of their identity could open them to possible retaliation. How can science best encourage experts to develop and share information that may reveal research misconduct without also encouraging the spread of meritless

accusations and personal attacks? Can journals and agencies do more to provide tools and information that speed the correction of the scientific record? One interesting experiment is PubMed Commons, a forum for postpublication peer review where commenters have to reveal their identities.[1]

A recent analysis of the role of social media and other nontraditional communications in several recent episodes in chemistry provides an optimistic view of the potential for these methods and tools to strengthen the self-correcting tendencies of science:

> The existence and vigorous participation of these forums in analyzing, challenging, and enhancing dialogue about the chemical literature and the human elements in research raise interesting questions with which the chemical community will have to grapple for the foreseeable future. Given the nature of transformational change over generations, it is also reasonable to predict that the younger generation which has grown up in the milieu of the breakthrough technology of the Internet will adapt and respond much more quickly to the changing norms of research and review discussed above. (Jogalekar, 2015)

## Investigating Misconduct and Taking Corrective Action

As discussed above, the U.S. federal research misconduct policy and its implementation in agency regulations place the primary responsibility for investigating research misconduct allegations on research institutions (HHS, 2005; NSF, 2002; OSTP, 2000). For extramural research funded by NSF and NIH, institutions are generally responsible for undertaking an initial inquiry into allegations to determine if a full investigation is warranted, to notify the agencies when such investigations are initiated, and to provide the agencies with the investigation reports, findings, and recommended actions when they are concluded for review. The U.S. federal policy specifies that a "preponderance of the evidence" standard be used to determine whether research misconduct has occurred, meaning it is necessary for 51 percent of the evidence to point toward misconduct in order to support a finding. The agencies evaluate the investigation reports, decide whether additional information is needed or not, and—in cases where they find that research misconduct has occurred—determine the remedies to be imposed.

NSF-OIG and ORI have several differences in how policies related to inquiries and investigations are implemented through their respective regulations. For example, NSF-OIG can perform inquiries and investigations itself when it chooses to or when an institution requests that it do so, since its authority comes from the Inspector General Act of 1978 (P.L. 95-452, 5 U.S.C. App.). ORI was created by the NIH Revitalization Act of 1993 (P.L. 103-43). ORI does not have the authority to perform its own investigations, although its staff assists institutions in their investigations and reviews the resulting reports. ORI may recom-

[1] (www.ncbi.nim.nih.gov/pubmedcommons/guidelines).

mend that HHS undertake its own investigation. HHS requires institutions that receive PHS funding to keep an assurance on file with ORI specifying that they have policies and procedures in place that comply with HHS regulations, and that they follow their own policies, or to file a Small Organization Statement if they lack the necessary resources to provide an assurance. Institutions also need to file an annual report to ORI to keep their assurances active. During 2011, 6,714 assurances were on file with ORI, including 425 from foreign institutions (ORI, 2012). NSF-OIG has no requirement similar to ORI's assurance program. There are also differences in the processes utilized for appealing research misconduct findings, with HHS specifying a more formal appeals framework than NSF. Details of NSF and HHS policies are contained in their implementing regulations, cited above, and on their websites.

Interactions between NSF-OIG, ORI, and institutions related to investigations go beyond formal oversight and reporting requirements. Both offices regularly send speakers to conferences and events to share information about their programs. In addition, ORI undertakes programs to train institutional research integrity officers (RIOs) and maintains a Rapid Response for Technical Assistance program to help institutions with advice, referrals, and assistance with forensic tools related to investigations. NSF-OIG can also provide advice and, as mentioned above, has the authority to undertake investigations itself.

As discussed in Chapter 5, the number of research misconduct inquiries and investigations has increased in recent years. For example, ORI received 423 allegations of research misconduct in 2012, far above the average of 198 received over the years 1992–2007 (ORI, 2012). A more recent annual total of 342 allegations for 2013 may indicate that the number of allegations is leveling off or even declining somewhat (ORI, 2014). For NSF-OIG, the number of allegations investigated grew from 45 in 2004 to 75 in 2014, and the number of research misconduct findings by NSF grew from 2 in 2004 to 20 in 2014 (NSF-OIG, 2015).

Information about the operation and performance of the inquiry and investigation systems overseen by NSF-OIG and ORI is available from several sources, such as the semiannual reports of NSF-OIG and the annual reports of ORI. NSF-OIG has made available a searchable database of case closeout memoranda, including memoranda from research misconduct cases and other types of cases that NSF-OIG investigates, such as financial fraud related to grants (NSF-OIG, 2015). ORI puts the summaries of completed cases that have resulted in findings of research misconduct on its webpage (ORI, 2015). Media reports of specific, notable cases are another source of information, but issues related to investigations may be covered and actual investigation reports may be released only when something has gone wrong with an institutional response. Only a limited amount of research has been done on institutional policies and capabilities.

In addition to these sources of information, this study benefited from briefings by agency officials (see Appendix B) and from responses to follow-up requests for information and clarification about specific issues. Undertaking a

comprehensive assessment of institutional and agency capabilities and performance related to research misconduct investigations would require a focused effort and access to a significant amount of information that is not currently available outside the institutions and agencies themselves. Nevertheless, it is possible to identify several issues where there is sufficient information to develop findings and recommend improved approaches, or at least to raise questions for future study and analysis.

### Different Approaches to Plagiarism

As noted in Chapter 4, the U.S. federal research misconduct policy defines plagiarism as "the appropriation of another person's ideas, processes, results, or words without giving appropriate credit" (OSTP, 2000). Differences between NSF-OIG and ORI in their approaches to plagiarism raise questions of whether the unified federal definition of misconduct is really "unified" and whether harmonization in the two approaches to implementation should be sought.

While both agencies state that they exclude "authorship disputes" as possible cases of misconduct, it appears that they draw the boundary between plagiarism and authorship disputes in different places. For example, ORI explains its policy as follows:

> Many allegations of plagiarism involve disputes among former collaborators who participated jointly in the development or conduct of a research project, but who subsequently went their separate ways and made independent use of the jointly developed concepts, methods, descriptive language, or other product of the joint effort. The ownership of the intellectual property in many such situations is seldom clear, and the collaborative history among the scientists often supports a presumption of implied consent to use the products of the collaboration by any of the former collaborators.
>
> For this reason, ORI considers many such disputes to be authorship or credit disputes rather than plagiarism. Such disputes are referred to PHS agencies and extramural institutions for resolution. (ORI, 1994)

The treatment of a case of apparent plagiarism from several years ago involving PHS-funded research raises questions about the implications of these differences in implementation. In 2011, postdoctoral fellow Heather Kling and her professor Karen Norris accused two other researchers at the University of Pittsburgh, Jay Kolls and Mingquan Zheng, of claiming credit for work that Kolls became aware of while serving on Kling's dissertation committee (Roth and Schackner, 2013). Kolls and Zheng applied for two federal grants and attempted to patent Kling and Norris's finding of a "vaccine against a lung disease known as pneumocystis," representing it as their own work (Roth, 2014). Arthur Levine, dean of the University of Pittsburgh School of Medicine, found Kolls and Zheng guilty of research misconduct; however, a faculty committee reduced

the finding to research impropriety, stating that it was "difficult to determine who first developed the idea" (Roth, 2014). Norris and Kling's lawsuit against Kolls and Zheng argued "that all the key lab work on the potential vaccine was carried out in the Norris lab," but that Koll's position as the head of a well-funded children's hospital may have played a role in the decision (Roth, 2014). Norris and Kling were added to the pending patent application, while Kolls and Zheng were subsequently removed (Roth, 2015). It is not clear whether or how the case was reported to ORI. Later communication between the university's research integrity officer and Norris indicated that ORI had "not taken an interest in the past in disagreements between investigators at the same institution" (Rosenberg, 2011). The university's Tenure and Academic Freedom Committee raised a number of concerns with how the allegations were handled (TAFC, 2013).

NSF-OIG appears to be more open to considering allegations of plagiarism against former collaborators than ORI is, including allegations involving collaborators with significant power differentials such as senior investigators and graduate students or postdoctoral fellows (see, e.g., NSF-OIG, 2013). In reviewing the closeout memos from NSF-OIG investigations dealing with allegations of "intellectual theft," it is clearly more difficult to establish that plagiarism has occurred in these cases than in plagiarism cases involving simple copying of text.

These apparent differences in policy implementation contribute to a different mix of case types handled by the agencies. NSF-OIG's largest category is plagiarism, while ORI's is data fabrication and falsification (Resnik, 2013). Contributing to this disparity is the fact that ORI handles significantly more fabrication and falsification allegations than NSF-OIG does. Changing ORI's approach to match that of NSF-OIG might have implications for the total number of cases handled by ORI. A more focused assessment of the two approaches, as well as those of other agencies, with access to more information than what was available to this committee would be needed to determine what specific changes are needed.

It is important to recognize the potential damage of maintaining a perception that a researcher can perhaps use the work of a student or another researcher at the same institution without permission or credit with near impunity while performing NIH-supported work—as apparently happened in the Kolls case described above—but would be investigated for misconduct if he or she took the same actions on NSF-funded work. Such inconsistency could contribute to a sense that norms and practices are not firm and clear. Again, a great deal of information is lacking, but the implications of the case are not encouraging.

### Institutional Capabilities and Performance

Since research institutions bear the primary responsibility for investigating research misconduct in the current U.S. system, their effectiveness in fulfilling this responsibility plays a significant part in determining how well the process of uncovering and investigating misconduct works overall. Effectively undertak-

ing inquiries and investigations includes a number of important elements, such as collecting and sequestering hard drives and other physical evidence, gaining necessary information from interviews with complainants, respondents, and others; observing confidentiality and due process protections for respondents; and ensuring that whistleblowers are not retaliated against. The Ryan Commission (discussed in Chapter 4) recommended that institutions have processes that are "accessible from multiple entry points," "overseen by individuals or by committees whose members are free from bias and conflict of interest," "based on independent investigation," "overseen by bodies that are separated in their investigatory and adjudicatory functions," "balanced in advocacy," "capable of preventing retaliation against participants," and "open" to the extent possible (Commission on Research Integrity, 1995). As noted above, the available information about institutional capabilities and performance is fairly limited. Still, some themes and lessons emerge from the information that is available.

Unevenness in institutional policies and capacity to investigate and address research misconduct allegations is an important challenge examined by the committee. As discussed in Chapter 4, institutions use a variety of definitions of research misconduct for internal purposes, even as they use the federal definition for the purpose of reporting misconduct to federal sponsors (Resnik et al., 2015). Differences in policies have been documented in other areas. For example, a 2000 survey of 156 institutions whose policies had been approved by ORI found that many institutional policies did not explicitly require researchers who encountered research misconduct to report it (CHPS Consulting, 2000).[2] A 2010 survey of medical schools and medical school researchers, a somewhat different target population from the 2000 survey, found that about one-third of the institutional policies do not explicitly require reporting of misconduct (Bonito et al., 2010). Many of the medical school policies do not contain clear guidance on the information that should be included in a research misconduct allegation. Most are clear about the particular institutional official or position that should receive the allegation. Almost all the medical school policies also have provisions for avoiding conflicts of interest, and most address the need to protect respondent and complainant rights.

A survey of medical researchers undertaken as part of the study of institutional policies showed that a majority are at least somewhat familiar with institutional and federal policies toward misconduct (Bonito et al., 2010). However, most made at least one error in going through a list of behaviors and identifying which ones constituted misconduct and which ones did not. Current information on the institutional policies for the full range of U.S. research institutions

[2] This is not to imply that the committee believes that this requirement should be in institutional policies. It was an issue of interest to ORI, and the responses illustrate institutional differences. Whether concerns are actually reported or not may have more to do with whether multiple entry points and other systems are in place to encourage reporting rather than a requirement of the institution's policy.

and how well those policies are understood by administrators, faculty, students, postdoctoral fellows, and others who could be affected would be helpful input to those working to assess and improve institutional performance.

Another salient aspect of an institution's capacity to investigate and address research misconduct allegations is the experience and ability of the institutional officials—including administrators as well as faculty members serving on investigation committees—responsible for implementing the institution's policies. Faculty investigation committees play a crucial role in overseeing investigations. At the same time, having competent and knowledgeable administrators is necessary to ensure that the committee has the necessary expertise and that other aspects of the investigation, such as evidence sequestration and documentation of interviews, are performed correctly. For some institutional officials tasked with these responsibilities but whose backgrounds and experience are primarily in research, this can pose a challenge. They may not have deep expertise in handling the complex administrative issues that can be encountered in research misconduct investigations (Gunsalus, 1998b).

A briefing by an ORI official during this study described the various ways that investigations can go wrong and provided anonymized examples (Garfinkel, 2012). For example, if relevant institutional personnel are not adequately trained regarding proper sequestration procedures for notebooks and data, sequestration of evidence may be inadequate or untimely. Institutional officials experience turnover and, given the low incidence of reported cases of misconduct, are rarely experienced at conducting the complex reviews required to resolve allegations of research misconduct. They generally carry myriad other duties, and research misconduct investigations can be very time-consuming and costly (Michalek et al., 2010). Institutional standing committees might not have the appropriate expertise to evaluate allegations in certain fields, leading to poor analysis and mistaken findings. Interviews may be poorly conducted or not be annotated. Investigation reports might fail to include sufficient evidence or rationale for findings. New allegations uncovered during the course of the investigation might not be followed up properly.

ORI also sponsored several surveys of research integrity officers aimed at learning more about the knowledge and preparedness of these institutional officials. In the first survey, the results of which were reported in 2009, RIOs were asked to complete an online survey recording the actions they would take in response to three scenarios that involved, respectively, sequestering evidence, protecting a researcher who had made allegations, and coordinating their own actions with those of the institutional review board (IRB) (Bonito et al., 2009). The responses were compared with model responses developed with two expert consultants. Several results of this survey were disquieting. For example, 97 percent of the respondents to the online survey identified fewer than half of the potentially appropriate actions for the three scenarios that had been given by the expert consultants. This indicates that a potentially significant proportion of RIOs

were not adequately prepared to fulfill their responsibilities. In addition, length of tenure in their positions was not positively correlated with greater knowledge, meaning that RIOs were not becoming more knowledgeable over time, on average. Having experienced specialized training, such as ORI's RIO "boot camp" seminars, was associated with greater knowledge. Another result of the survey was that less than one-fifth of respondents were formally designated as an RIO or compliance officer by their institutions. Many of those who were performing the functions of a RIO had other responsibilities in areas such as grants management.

A second survey of RIOs commissioned by ORI focused on how the RIOs interacted with those making allegations of misconduct (Greene et al., 2011). ORI originally wanted to survey those who had made allegations but found that it is impossible to identify, locate, and survey the complainants of closed cases due to the current interpretation of regulations protecting confidentiality. The survey involved interviews with 102 RIOs. They were asked whether they discussed four key topics with complainants: "the resolution process, anonymity and confidentiality, institutional responsibilities, and potential adverse consequences" of coming forward with an allegation. Less than half reported discussing all four topics with those considering making an allegation. The report pointed out that it would be helpful for RIOs to discuss this issue with potential complainants because many of those who come forward with research misconduct allegations have reported experiencing retaliation or other adverse consequences (Lubalin et al., 1995). One recommendation made in the 2011 report of the survey is that ORI encourage RIOs to use a prepared script or other memory aid to ensure that all the topics are covered.

Another problem that arises in some research misconduct cases and their handling by institutions is delay in reaching the inquiry and investigation stages. As noted earlier in this chapter, a 2013 analysis of 120 well-known cases of research misconduct found that there was a failed attempt to report misconduct in 28 percent of the cases (DuBois et al., 2013b). Publicly available details of several notable cases of research misconduct provide additional insights. For example, in the translational omics case at Duke University, the system failed at several points (see Appendix D). At the laboratory level, Joseph Nevins did not thoroughly check the data reported by Anil Potti until Potti's misrepresentations in his resume were uncovered and publicized, even though the data had been questioned by experts over the course of several years (CBS News, 2012). At the institutional level, the department did not perform a thorough audit of the data after a graduate student raised concerns and asked that his name be taken off articles that would be submitted based on the work. This graduate student was assigned to another lab. When Duke ultimately launched an investigation sometime later, the external investigation committee was not given all the relevant information, a circumstance that was probably at least partly responsible for the committee recommending that clinical trials based on the work should continue.

Two factors present in the Duke case are sometimes seen in other cases

where the launch of an inquiry or investigation is delayed. First, at the outset, the concerns and evidence of internal and outside researchers were not brought forward as formal allegations of misconduct but, rather, as concerns and questions about possible errors. Second, the researcher whose work was being questioned was closely associated with a high-prestige researcher. As is seen in other contexts such as financial or political misconduct, officials may have biases that filter how they hear concerns or lead to reluctance to make or aggressively pursue allegations of wrongdoing against powerful people in their own organization or against people closely associated with them. The researchers raising the concerns or questions may hesitate to move forward to a formal allegation, and the absence of an allegation may override for a time the suspicions of institutional officials based on an impartial assessment of the evidence. The path of least resistance might be to continue to delay action.

Additional information about how institutions address research integrity issues more broadly has emerged via administration of the Survey of Organizational Research Climate (SOURCE), reported as part of the Project on Scholarly Integrity (PSI) undertaken by the Council of Graduate Schools and in other contexts (CGS, 2012). A Research Integrity Inventory Survey was also administered as part of PSI.[3] SOURCE was developed partly in response to the recommendations contained in the report *Integrity in Scientific Research* (IOM-NRC, 2002; Thrush et al., 2007). SOURCE's questions focus on institutional resources to foster responsible conduct, policies and regulations, subunit (i.e., departmental) norms, advisor-advisee relations, and integrity inhibitors and expectations (Martinson et al., 2013). The final report of the PSI gave aggregated results of the six institutions that administered SOURCE as part of the project, and a subsequent article authored by several of the participants in the PSI has reported more detailed results for a subset of three of the participating institutions (Wells et al., 2014). SOURCE indices have also been shown to correlate with a broad range of research-related behaviors (Crain et al., 2013). SOURCE is available to institutions that wish to utilize it, and institutions can also contract with Ethics CORE (nationalethicscenter.org/sorc) at the University of Illinois to administer it and compile the data.

The coverage of institutional investigations also affects the ability of journals to correct the research record and of sponsors to take corrective action. Do institutions have an obligation to investigate a researcher's work beyond the specific publications or proposals that are subject to allegations of research misconduct? In an international example, the three Dutch universities where Diederik Stapel was educated and employed came together and investigated all the work that he produced in his career, from his PhD dissertation onward (Levelt et al., 2012). The resulting report, which was published in its entirety and translated into

[3]Additional information on these resources is available at www.scholarlyintegrity.org/ShowContent.aspx?id=400#.

English, is a significant contribution to social psychology and to the broader understanding of research misconduct. In other cases, institutions may stick to investigating the work that is subject to allegations due to resource constraints, barriers to involving other institutions, or other reasons.

One important conclusion from this discussion of institutional capabilities and performance is that significant gaps exist in the information available to institutions as well as to the rest of the research enterprise about how allegations are handled, what challenges arise, and how successful institutions are able to ensure effective performance. Several items in the institutional best practices checklist discussed in Chapter 9 are aimed at filling in this information deficit at the institutional level. The occasional surveys supported by ORI have shed light on important aspects of institutional responses, but additional research to assess aggregate trends and needs could yield valuable insights that would enable the entire system of investigating research misconduct in the United States to operate more effectively.

### Taking Corrective Actions

Corrective actions may be taken in response to research misconduct findings, and these take several forms. The employing institution should notify journals that have published articles based on fabricated or falsified data so that they can be retracted. The institution will determine whether to send a letter of reprimand to the guilty researcher, suspend him or her, or terminate employment. Federal actions can be taken if the research in question was performed with federal government support. An agency can suspend or terminate the award, institute requirements that the researcher's actions be supervised, or debar the researcher from receiving support or from participating in agency review or advisory activities in perpetuity or for a period of time (OSTP, 2000).

Both NSF-OIG and ORI processes have avenues for appeal available, and when these are exhausted, accused researchers can go to court. Several examples reported in the press in recent years involved researchers who appealed research misconduct findings and had some measure of success in having penalties overturned or reduced (Cossins, 2013; Kuta, 2014).

As noted above, agencies differ in their implementation of federal policy. ORI's policies and procedures in handling investigations generally involve more formal requirements than do those of NSF-OIG. For example, ORI publishes the names of all researchers found guilty of research misconduct on its website. NSF-OIG only allows the public to ascertain whether researchers have been debarred or suspended—about 25 percent of NSF's cases—and the names are not published. The names of those who have been debarred or suspended are entered into the System for Award Management, a public database (www.sam.gov/). Although the database is searchable, one needs to enter a name to perform a search, so it is a useful tool for universities that might be hiring a faculty member or an

agency checking on a grant applicant to determine if he or she has been debarred or suspended. However, the database will not generate a list of names of researchers who have been found guilty of misconduct, and the entries do not mention the reason for debarment or suspension. Researchers can be debarred for misappropriating award funds and for other causes in addition to research misconduct.

One consequence of this difference in implementation between agencies is that a researcher who is found guilty of misconduct when performing NSF-funded research is unlikely to suffer from public disclosure during that researcher's subsequent professional life unless the case was reported in the media, while an NIH-funded offender will certainly be exposed. Since the disclosure itself represents a significant consequence—perhaps the most significant consequence—this difference in policy implementation between NSF-OIG and ORI in fact constitutes a clear disparity in the severity of corrective actions (see discussion of the relatively new Department of Veterans Affairs policy below). Research exploring the long-term consequences of being found guilty of misconduct and having that judgment publicly disclosed found that while many offenders left research, 43 percent of those who had been in academia and could be traced still had academic research jobs some years later (Redman and Merz, 2008). Some efforts have been made to develop educational programs aimed at rehabilitating researchers who have been accused of misconduct (Cressey, 2013; DuBois et al., 2016).

The continued existence of this disparity is problematic both for individual researchers, who are held to different standards of accountability based on their sources of funding, and for the research institutions employing the researchers, which must implement policies for such a heterogeneous aggregate of researchers. A federal effort to bring about greater consistency between agencies in the implementation of the research misconduct policy could address this issue, as could other initiatives such as reconciling differences in the handling of plagiarism allegations. However, it is not obvious how implementation should be made consistent. The appropriate approach might depend on what one sees as the ultimate goals of corrective action. Both approaches—public reporting and maintaining anonymity—have positive and negative aspects (Parrish, 2005). Should those found guilty of research misconduct have their research careers ended, or are there some cases where errant researchers can be rehabilitated? Should younger researchers be treated differently from those with more experience? What are the risks to future research and the potential damage to future collaborators in cases where the identity of those found guilty is not disclosed? Is it possible to craft an approach where those found guilty of the most egregious offenses are exposed, while those whose misconduct is less consequential, particularly younger researchers, are not? Policy makers and members of the research enterprise differ on these questions. One provision of the America COMPETES Reauthorization Act passed by the House of Representatives in 2015 would have required NSF-OIG to make the names of "principal investigators" public in cases of misconduct, effectively harmonizing implementation around ORI's current

practices. This provision was not included in the version of the bill that was ultimately passed by both houses of Congress and signed by President Obama in 2016 (American Innovation and Competitiveness Act of 2016).

A complicating factor in efforts to harmonize the approaches of federal agencies is that institutions have different policies and approaches to identifying employees who commit various types of offenses, including research misconduct. For example, some institutions do not normally publicize the fact that researchers have been found guilty of misconduct, and may make investigation reports publicly available only in response to Freedom of Information Act requests. The University of Kansas has taken a different approach, occasionally publishing "public censure" items in its employee newsletter in response to cases of research misconduct and other prohibited behavior (University of Kansas, 2013).

In addition to the administrative actions that can be taken by institutions and research funding agencies, researchers who commit misconduct can face criminal prosecution under certain circumstances. For example, in 2015, Iowa State University researcher Dong-Pyou Han was prosecuted and convicted for fabricating and falsifying data in HIV vaccine trials (Reardon, 2015). The prosecution occurred after the institution had completed its investigation and ORI had issued its findings and administrative actions, and after Iowa Senator Charles Grassley called attention to the case. Han received a sentence of 57 months in prison and $7.2 million in fines. Over the years, several other notable cases of misconduct have led to prosecutions, including that of anesthesiologist Scott Reuben (discussed above), although these cases are unusual (Bornemann-Cimenti et al., 2015). Decisions on whether to prosecute such cases will depend on the likelihood of success and how possible misconduct cases rank versus other potential uses of available prosecutorial resources. It is important to note that the standard of proof in criminal prosecutions—proof "beyond a reasonable doubt"—is significantly higher than the "preponderance of the evidence" standard that prevails in institutional research misconduct investigations and federal oversight of these investigations. In recent years, there have been calls for research misconduct to be treated as a crime more frequently than it has been up to now (Smith, 2013).

Finally, researchers who commit misconduct may face civil liability, and institutions may face civil penalties if they are negligent in their oversight or responses. One avenue for pursuing such penalties is the False Claims Act, which allows the federal government to recover damages and penalties from those who make false claims on the government (Kalb and Koehler, 2002).

### Research Misconduct and Other Regulatory Frameworks

The implementation of policies to address research misconduct by federal agencies and institutions can sometimes be intertwined with and affected by other regulatory frameworks that govern certain types or aspects of research. Although reviewing these issues and related evidence contributes to an understanding of

some of the tasks and challenges facing agencies and institutions, developing solutions or new approaches is largely outside the scope of this study.

The most obvious example of regulatory intertwining concerns the regulations designed to protect the human subjects of research in clinical trials and other settings. The basis of federal policy in the area of human subjects protection is the "Common Rule," covered in 45 CFR Part 46, of the Code of Federal Regulations, "Protection of Human Subjects," which covers research supported by federal agencies or subject to federal regulation, such as privately funded clinical trials that are subject to oversight by FDA. Institutions performing research on human subjects are required to undertake a prospective ethical review of proposed research through a standing institutional review board or other mechanism, ensure that human research subjects provide "informed consent" to participation in the research, and promptly report any unanticipated risks or failure to comply with regulations during the course of the research.

A 2014 report outlines the differences between the research misconduct and human subjects protection regulations and explains the complexities and challenges that can arise for institutions as they seek to comply with both (Bierer and Barnes, 2014). For example, fabrication or falsification of data in a federally funded research project involving human subjects may trigger fact-finding and enforcement processes under both sets of regulations. In general, the research misconduct regulations are more detailed and specific for investigation procedures (including opportunities for appeal), confidentiality requirements, standards of proof, and other issues than are the human subjects protection regulations. Examples of issues and questions that may arise in cases that fall under both sets of regulations include how to provide an IRB with access to data that have been sequestered as part of a research misconduct investigation and what weight (if any) a research misconduct investigation should give to an IRB finding that allegations of noncompliance with Common Rule standards have not been substantiated (Bierer and Barnes, 2014).

Additional issues arise in connection with reporting and information flows between officials working to ensure human subjects protection and those responsible for investigating research misconduct allegations. As illustrated by the discussion above of clinical trials where an FDA investigation found significant problems such as falsification of data, but where the research was published with no indication of a problem and there is no record of a research misconduct finding, there appear to be shortcomings in how information flows between the two regulatory and compliance systems.

Another area where human subjects protection regulations overlap with research misconduct regulations is in education, since human subjects protection is included as one of the nine core areas of responsible conduct of research (RCR) education as defined by NIH (NIH, 2009). RCR education is discussed further in Chapter 9.

Starting in 2011, the Department of Health and Human Services embarked

on a process of revising the human subjects protection regulations that was in process during most of the time when this study was being undertaken (HHS, 2011a). In 2015 a Notice of Proposed Rulemaking was published describing the major changes suggested to the Federal Policy for the Protection of Human Subjects. The Notice of Proposed Rulemaking proposes changes to the rules regarding informed consent, including revisions to consent forms and research subjects providing "broad" consent for secondary research, and to the oversight system, through "making the level of review more proportional to the seriousness of the harm or danger to be avoided" (HHS, 2016). It is unclear how the resulting changes will affect agencies and institutions as they seek to manage areas of overlap between the research misconduct and human subjects regulatory frameworks.

Another area of regulation that has some relationship with research misconduct policies involves the requirements for disclosing and managing possible financial conflicts of interest in research. Conflicts-of-interest reporting is an issue currently in flux, with many differences between existing policies, which are not uniform between agencies, and compliance generally does not raise issues of overlap with research misconduct policy. In response to research showing that conflicts-of-interest reporting can have perverse effects by providing a "strategic reason and moral license" to exaggerate advice, policies, and regulations aimed at ensuring that financial conflicts of interest do not adversely affect research or skew results have been changing in recent years, and the impacts on how research misconduct is addressed may change in the future (Cain et al., 2005, 2011; Koch and Schmidt, 2010; Loewenstein et al., 2011). For example, HHS revised its policies toward financial conflicts of interest in PHS-funded research in 2011, which changed some of the reporting requirements of researchers and institutions (NIH, 2011). In NSF-funded research, institutions are required to certify as part of the proposal process that they have a policy covering conflicts of interest and that the proposed research complies with that policy (NSF, 2014). The National Academies' report *Optimizing the Nation's Investment in Academic Research* recommends a federal government–wide financial conflicts-of-interest policy that differentiates "requirements for financial interest disclosure and management for research that does and does not involve human subjects" in an effort to reduce the time and cost burdens of multiple existing policies (National Academies of Sciences, Engineering, and Medicine, 2016). As discussed in Chapter 4, some countries treat failure to disclose potential conflicts of interest as a form of research misconduct. Conflicts-of-interest regulations are largely outside the scope of this study and so, in this report, addressing potential conflicts of interest in research is treated as an issue to be addressed through best practices, as discussed in Chapter 9.

### Correcting the Research Record: Journals and Retractions

One important aspect of addressing research misconduct is correcting the published research record through the retraction of journal articles. Retractions are discussed in several other places in this report, including Chapter 5, which discusses the significant increase in retractions that has occurred in recent years and the extent to which statistics on retractions are a useful proxy for the incidence of misconduct (Grieneisen and Zhang, 2012). Retractions, while not rare at this point, are still relatively unusual. Here, it is important to note that retracting articles is not always a consistent or straightforward process and to identify issues that might be addressed by journals or by other stakeholders.

One way to gain insight into retractions is to examine a case where a researcher was found to have committed misconduct and where a number of his or her articles needed to be reanalyzed and possibly retracted. An analysis of how journals responded following a finding of research misconduct against University of Vermont obesity and aging researcher Eric Poehlman is one such case (Sox and Rennie, 2006). In this case, 4 of the 10 articles identified by ORI as being based on fabricated or falsified data had not been retracted more than a decade after the finding of misconduct (McCook, 2015). There is a mix of reasons for why individual papers have not been retracted, with several having only been subject to a correction notice.

Although journals are not equipped to investigate allegations of research misconduct, they may have strong evidence of misconduct developed through the use of software that detects image manipulation or through other technological tools. In the absence of a finding of misconduct or a request by an institution to retract an article, a journal might hesitate to move forward. Some retractions "can involve unavoidable delays of years because of some combination of the complexity of the science, disputes between co-authors, the need to await outcomes of lengthy investigations, and disputes over these proceedings" (*Nature*, 2014). In the absence of an institutional finding, a journal may be concerned that citing misconduct as the cause of a retraction would open the door to a libel suit or other legal action, although it is unclear if such legal action has ever been taken by an author (Wager, 2015b).

Another issue that arises with retractions is that retracted work may continue to be cited. For example, a recent analysis of 25 retracted papers by Scott Reuben found that, 5 years after the retractions, nearly half the papers were still being cited, with most of the citations not mentioning that the work had been retracted (Bornemann-Cimenti et al., 2015).

Chapter 9 discusses best practices that should be adopted by journals in the area of retractions. Technological tools that allow researchers to identify the publisher-maintained version of an article and the development of master databases of retractions will likely reduce the phenomenon of retracted work being cited in the future.

## Other Issues, Gaps, and Inconsistencies

### Privately Funded and International Research

As mentioned in the Chapter 4 discussion of research misconduct definitions, the federal research misconduct policy only applies to federally funded research. The federal requirement that institutions report investigations and their results to funders does not apply to privately funded research, including research supported by international sponsors. Institutional policies do not make a distinction between funding sources in how allegations should be handled, and there appears to be no evidence indicating whether institutions make such distinctions in practice or not. The results of these investigations may not be made public, so it is not possible to track incidence or trends at the aggregate level. However, some cases of misconduct where work needs to be retracted do become publicly known.

There are several notable cases where misconduct in privately funded research has been investigated and addressed. One example is the data fabrication and falsification by Jan Hendrik Schön of Bell Laboratories (Goodstein, 2010). Results of his seemingly groundbreaking research on semiconductor materials were published in a number of prestigious journals, mainly between 2000 and 2002. In early 2002, other researchers within Bell Labs and outside began to raise questions about Schön's work, and Bell Labs set up a committee to investigate. The committee released its report later that year, finding that Schön had committed scientific misconduct (Beasley et al., 2002). The report served as the basis for retraction of numerous papers. The committee stated that there was no evidence that any of Schön's coauthors were aware of or involved with the misconduct, noting that in only a few cases had coauthors had any involvement in fabricating the devices in question, designing or performing the experiments, observing the reported phenomena, or collecting or analyzing the data. The Schön case raises the question of whether coauthors bear responsibility for reviewing or confirming the work of their collaborators; this issue has appeared in several high-profile cases since that time, such as the stem cell case of Hwang Woo-suk (see Appendix D).

### Sabotage as Falsification

Chapter 4 contains a discussion of whether cases where researchers sabotage the experiments of others or abscond with vital data should be considered research misconduct. In at least one case, ORI has treated sabotage of experiments as research misconduct. In a case from several years ago, Vipul Bhrigu, a postdoc at the University of Michigan Medical School, was found to have tampered with the experiments of Heather Ames, a graduate student in his lab, which caused false results to be reported in the research record (HHS, 2014a; Maher, 2010). The tampering had been videotaped. Bhrigu also was convicted of malicious destruction of personal property (Maher, 2010).

In another incident reported in the media, Polloneal Jymmiel Ocbina, a postdoc at Yale, was videotaped tampering with the zebrafish experiments of another postdoc, Magdalena Koziol, and left Yale without being charged with a crime (Enserink, 2014). Koziol later sued Yale and her supervisor, Antonio Giraldez, for not allowing her to speak about the case to sponsors in explaining why she had not made more progress in her work.

It is well established that tampering with data and experiments to obtain false-positive results constitutes falsification. Given that the Bhrigu case has established a precedent and conditions under which tampering to cause another researcher to obtain false-negative results also constitutes falsification, a useful way to ensure greater consistency in federal agency implementation of the research misconduct policy might be to examine how institutions treat cases of sabotaging experiments and absconding with data, perhaps through a survey or other mechanism.

### Issues Raised by the Policies and Practices of Federal Agencies Other Than ORI and NSF-OIG

Much of the discussion of policies and policy issues in this report focuses on ORI and NSF-OIG, which oversee the handling of research integrity issues by grantees of the Department of Health and Human Services (the bulk of these being grantees of the National Institutes of Health) and the National Science Foundation, respectively. Looking at a few statistics shows why these agencies have a disproportionate importance in the implementation of federal research misconduct policy. NSF and HHS account for about 80 percent of the federal research and development funding that is provided to academic and private nonprofit organizations (NSB, 2014b), and authors affiliated with academic and private nonprofit organizations account for about 80 percent of the research articles published by U.S. authors, with authors affiliated with industry, federal agencies, and federally funded research and development centers accounting for most of the rest (NSB, 2016). NSF and HHS clearly play leading roles in federal support for research that results in published articles.

Despite the lack of federal government-wide statistics or reporting on research misconduct investigations and findings, the available evidence indicates that NSF-OIG and ORI account for the vast majority of total activity. Also, federal agencies other than ORI and NSF-OIG do not appear to produce regular public reports on how many investigations have been launched and their resolution, as ORI and NSF-OIG do. As described below, agencies follow a variety of approaches toward making information about research misconduct investigations public.

Despite the understandable focus on NSF-OIG and ORI, other federal agencies that perform and/or support research are also obliged to implement the federal research misconduct policy. These agencies may face different challenges

depending on whether their research programs are mainly intramural or extramural and on other factors. Also, in some cases the handling and resolution of misconduct allegations affecting research supported or performed by other agencies have led to questions or controversy. Although a detailed review of how all agencies are implementing the research misconduct policy is beyond the scope of the study, examining several examples serves to illustrate that efforts to assess and improve performance by federal agencies would contribute to fostering research integrity within the federal government and beyond.

Part of the context is the scientific integrity initiative that the Obama administration undertook during its first term, described in Chapter 3. Executive branch agencies were instructed to develop policies to ensure the credibility of government research and prevent bias in how science is used in policy making. As part of the initiative, some agencies reviewed their existing policies. For example, a 2010 review at the Department of the Interior (DOI) found no comprehensive scientific integrity policy at the department level, although an effort to develop a policy and code of conduct to implement the 2000 federal research misconduct policy had been attempted and failed (DOI, 2010). In 2007, one of the DOI's constituent agencies, the U.S. Geological Survey, issued a scientific integrity policy that implemented the 2000 federal research misconduct policy. Following up on the 2010 review, DOI developed a comprehensive department-wide policy that was implemented in 2011 and updated in 2014 (DOI, 2014).

DOI agencies such as the U.S. Geological Survey and the Fish and Wildlife Service both perform intramural research and support extramural research. Investigations of possible research misconduct and other breaches of scientific integrity are overseen by scientific integrity officers appointed by DOI's constituent agencies. DOI also posts summary results of the research misconduct investigations that it has undertaken and concluded since 2011 (DOI, 2015). The most controversial DOI scientific integrity cases of recent years have revolved around establishing and reporting the scientific basis for agency policies and positions, rather than fabrication, falsification, and plagiarism. For example, disputes have emerged, and investigations of alleged breaches in integrity have been undertaken, over the development and presentation of the scientific evidence used to predict the impacts of such actions as building the proposed Keystone XL pipeline and removing dams from the Klamath River. A case of data falsification at a USGS laboratory that the agency investigated and confirmed, as described in a later report of the DOI Office of Inspector General, illustrates that research misconduct may occur in research performed at government laboratories (DOI-OIG, 2016).

The Department of Defense (DOD) is an important performer and sponsor of research. DOD issued a directive in 2004 that delegated to component agencies the responsibility for developing and implementing procedures to foster research integrity, including procedures for addressing allegations of research misconduct (DOD, 2004). The DOD directive also defines standards and requirements for

those procedures, referring to the definitions set out in the 2000 federal policy. In response to the federal scientific integrity initiative of 2010, DOD developed a separate policy that covers the utilization of science in policy making, media relations, and other issues distinct from addressing research misconduct.

A research misconduct investigation concluded in 2007 shows that challenging issues may arise in connection with addressing misconduct allegations in DOD-sponsored research (Godfrey, 2007). In that case, an engineering team from the Massachusetts Institute of Technology's Lincoln Laboratory that evaluated a 1998 ballistic missile defense flight test was accused of research misconduct. The investigation was delayed for several years when DOD refused to allow access to classified information deemed essential to undertaking the investigation. Once access was granted, the investigation proceeded, resulting in a finding that research misconduct had not occurred and exoneration of the Lincoln Lab authors, Ming-Jer Tsai and Charles Meins (Godfrey, 2007). In addition to summarizing the investigation, the final report contains suggestions for improvements in conducting future investigations.

The Department of Veterans Affairs (VA) undertakes a large program of clinical and discovery research programs, budgeted at $1.8 billion for fiscal 2015, combining the VA's own dedicated research budget, medical care support, other federal resources, and nonfederal resources (VA, 2015). The VA's Program for Research Integrity Development and Education oversees training and credentialing in areas related to human subjects protection. The VA also has detailed policies and procedures for dealing with research misconduct allegations, with the most recent version being issued in early 2014 (VA, 2014). These policies and procedures were reviewed and revised prior to being reissued, with a number of substantive changes introduced to clarify roles and improve procedures for conducting inquiries and investigations and to harmonize VA's policies with those of the Public Health Service that are implemented by ORI (Bannerman, 2014).

Research integrity officers are appointed at all VA facilities with an active research program. Depending on how the processes of conducting the initial inquiry, undertaking the investigation, reviewing the report, adjudicating the result, and overseeing any appeal proceed in specific cases, there are defined roles for the director of the facility where misconduct has been alleged, the director of the Veterans Integrated Service Network (VISN) that includes the facility, the VA Office of Research Oversight, the VA Office of General Counsel, and the VA Under Secretary for Health.

One interesting aspect of the revised VA research misconduct policy is that it includes specific provisions for publication of findings, which the previous policy lacked:

> **Publication of Final Findings of Research Misconduct.** For all findings of research misconduct adjudicated by a VISN Director and upheld by the Under Secretary for Health on appeal, if any, VA may publish the respondent's name, the respondent's current or former VA position, a detailed summary of the find-

> ings, and the corrective actions imposed, in any venue deemed appropriate. Such venues include, but are not limited to, Government exclusionary lists (if relevant), the Federal Register, ORO's Web site, other VA publications, and media outlets. VA may also provide the information referenced in this paragraph to the respondent's current employer and academic affiliates, as well as other entities whose notification would be necessary to implement a corrective action (e.g., journal editorial boards). ***NOTE:*** *In those cases where there is a determination that the extent of the research misconduct is significant and/or the possible or actual consequences of the research misconduct are significant, it is considered to be in the interests of both VA and the scientific community to publish final findings of research misconduct*. (VA, 2014)

This approach to publishing investigation results differs from those of NSF-OIG and ORI discussed above. The policy allows, but does not require, the names, findings, and corrective actions related to misconduct to be published, preserving discretion for the agency.

The U.S. Department of Energy (DOE) is also a significant sponsor of research. Much of the research that DOE supports is performed at its National Laboratories and user facilities, most of which are managed and operated by contractors. DOE's research misconduct policy was adopted in 2005 and specifies that research misconduct allegations should be referred to "the DOE Element responsible for the contract or financial assistance agreement" (10 CFR Parts 600 and 733; 48 CFR Parts 935, 952, and 970). The policy also specifies that the DOE element in question should consult with the DOE Office of the Inspector General (OIG), which can decide whether to investigate the allegation itself. If DOE-OIG declines to investigate, the allegation is referred to the contractor or grantee. The requirements for contractors and grantees regarding research misconduct investigations are covered in more detail in DOE's contracting regulations (48 CFR Chapter 9). The contractor or grantee is primarily responsible for adjudication and determination of corrective actions, although DOE reserves the right to take additional action.

Questions about DOE's policies were raised in connection with an investigation of an anonymous allegation against a research group at Oak Ridge National Laboratory (Reich, 2011). In that case, the lab's investigation found that Stephen Pennycook's group had not fabricated or falsified data. In the aftermath of this case, questions were raised in a Freedom of Information Act lawsuit by *Nature* reporter Eugenie Samuel Reich about whether DOE's oversight of research misconduct investigations by contractors and grantees was adequate, and whether DOE should consider establishing a new organization focused on performing such oversight (Reich, 2011). A 2014 audit of DOE's management of research misconduct investigations reported that around 30 research misconduct allegations had been received by DOE's Office of Science and the National Laboratories between 2009 and 2013 (DOE-OIG, 2014). It is unclear how many of these allegations proceeded from the inquiry stage to an investigation. DOE-OIG audited

the responses to 21 allegations and found that they were addressed appropriately. However, DOE-OIG found several cases where requirements to report allegations to the OIG or the contracting officer were not followed and others where contractors did not follow their own misconduct investigation procedures. The report recommended that DOE's Office of Science "provide additional education and guidance on the procedures and responsibilities for conducting research misconduct allegation reviews to Department officials, laboratories, and financial assistance recipients" (DOE-OIG, 2014).

The Environmental Protection Agency (EPA) also performs research and supports extramural work. Its policy on addressing research misconduct allegations made against EPA employees and contractors was adopted in 2003 and specifies investigative and reporting requirements (EPA, 2006). As is the case in several of the other agency examples, the agency's Office of the Inspector General has an important role in overseeing responses to allegations, including the authority to step in and undertake an investigation under certain circumstances. Research misconduct is also discussed in EPA's scientific integrity policy, which was adopted in 2014 (EPA, 2012). This newer policy does not replace or amend the procedures for responding to allegations but does identify a new position within EPA, the Scientific Integrity Official, who is responsible for working to promote scientific integrity within EPA.

## ADDRESSING DETRIMENTAL RESEARCH PRACTICES

The concept of detrimental research practices and specific examples of DRPs are discussed in Chapter 4. Chapter 5 describes the negative impacts of DRPs on the research enterprise in terms of misallocated financial resources and wasted effort. The sum total of these negative impacts may be greater than the harm done by research misconduct. Some detrimental research practices related to authorship that do not constitute misconduct, such as honorary authorship, are discussed below in a section focused on authorship issues and challenges.

The discussion in Chapter 5 also explains how some detrimental research practices, such as misleading statistical analysis that falls short of falsification, incomplete reporting of results that leads to misrepresentation of findings, and the failure to retain or share data and other information (such as code) underlying reported results, are implicated in the reproducibility problem—that an alarmingly high percentage of the reported findings in certain fields cannot be replicated. The example of several specific cases also shows that DRPs are closely connected with research misconduct. Tolerance of DRPs in certain fields, as embodied in the policies of journals and sponsors, as well as in accepted practices at the laboratory level, can delay or prevent the discovery of misconduct.

To the extent that standards can be improved and tolerance for DRPs can be lowered or eliminated, fabrication and falsification of data will be more easily and quickly uncovered in many cases. In addition to improving the efficiency of

research in these fields in the production of reliable knowledge, the development and implementation of higher standards and improved practices will make it more difficult for long careers to be built on fraudulent work, as Stapel and Reuben were able to do (as described in Chapter 5). We can expect some, perhaps many, researchers inclined in that direction to be deterred. Discouraging, reducing, and eliminating DRPs will support and strengthen the effective operation of science's self-correcting tendencies.

An example from high-energy physics illustrates the value of good research practices in the process of reporting results, identifying and correcting errors, and confirming findings. The apparent discovery in 2002 of pentaquarks, a short-lived particle made up of five subatomic quarks, quickly led to a number of confirmatory reports (Chalmers, 2015). Previous theoretical work had predicted the existence of pentaquark states. However, subsequent efforts to replicate these results at a higher level of sensitivity failed and appeared to prove that pentaquarks do not exist. In the most recent development, researchers analyzing data collected from an experiment at the Large Hadron Collider at CERN appear to have confirmed the existence of pentaquarks (Chalmers, 2015). This episode shows the value of reporting results and the underlying information so that others can confirm results and extend the findings, and serves as a reminder that science often proceeds through various twists and turns in the accumulation of reliable knowledge.

A widely reported 2011 article claiming that bacteria could grow without phosphorus by using arsenic instead is an example showing the value of postpublication community review in identifying problems with work that are unrelated to misconduct (Wolfe-Simon et al., 2011). The article was criticized immediately and refuted by later work (Kaufman, 2012).

However, as discussed in Chapter 5, current standards and practices in particular fields may not be adequate to counteract widespread lack of rigor in study design, bias in selecting data or publishing results, and other errors. Developing appropriately high standards in research and ensuring their wide adoption are complex tasks requiring the contributions of various stakeholders with different perspectives and incentives. The heightened attention that the reproducibility problem has recently attracted provides an opportunity to make progress.

Better awareness and recognition that there is a problem at the level of specific fields and disciplines, and communication of this awareness to institutions and investigators, can be an important starting point. A well-designed replication effort can provide insights on the nature and possible scope of problems. A recently published effort to reproduce 100 studies published in three psychology journals is a valuable demonstration along these lines (OSC, 2015). The replication effort was undertaken as an open, global collaborative and involved contacting the original authors for materials and asking them to review the replication study protocol, public registration of the protocol, and public archiving of the replication materials and data (Aarts et al., 2015). The result was that 36

percent of the replication efforts yielded significant results versus 97 percent of the original studies. In addition, the effects found in the replications averaged half the magnitude of the originals. The effort also found that the original results from cognitive psychology were more robust than those from social psychology. While pointing out some caveats and uncertainties in interpreting the results, the summary of the replication effort yielded important insights into irreproducibility in psychology and its likely sources:

> More generally, there are indications of cultural practices in scientific communication that may be responsible for the observed results. Low-power research designs combined with publication bias favoring positive results together produce a literature with upwardly biased effect sizes. This anticipates that replication effect sizes would be smaller than original studies on a routine basis—not because of differences in implementation but because the original study effect sizes are affected by publication and reporting bias, and the replications are not. Consistent with this expectation, most replication effects were smaller than original results, and reproducibility success was correlated with indicators of the strength of initial evidence, such as lower original *P* values and larger effect sizes. This suggests publication, selection, and reporting biases as plausible explanations for the difference between original and replication effects. The replication studies significantly reduced these biases because replication preregistration and pre-analysis plans ensured confirmatory tests and reporting of all results. (OSC, 2015)

### Strengthening Standards and Ensuring Transparency

Detrimental research practices and some amount of failure to reproduce research results are not new problems. When the research enterprise was smaller and researchers in specific fields were more likely to know each other, personal communications about irreproducible work could be shared privately (Begley and Ioannidis, 2015). This undoubtedly still occurs, although this informal knowledge that certain work is unreliable may not be widely shared. As the enterprise has grown larger and competition has become more intense, the incentive to publish more articles has become stronger. In some of the specific examples described in this report, there appeared to be little or no checking of data at the laboratory or institutional levels, raising the question of whether ineffective supervision is widespread in certain fields and institutions. Funders and journals may not insist that researchers make data, code, and other information underlying results available. These factors, in combination, may create environments where publication bias and selection bias can go relatively unchecked and influence reported work.

Another important point discussed in Chapter 5 is that some false results will and should continue to appear in the normal course of science. Introducing practices aimed at reducing the irreproducibility rate to zero across all fields would be counterproductive and impose significant costs.

Clearly, improving transparency is a key factor in making improvements.

Chapter 8 is devoted to a discussion of best practices for researchers, research institutions, journals, sponsors, and societies. Much of the best practices discussion is related to improving transparency. Broad principles related to transparency in such areas as sharing data should be observed as widely as possible across all fields; these principles are the focus of several recommendations in Chapter 10. As discussed in Chapter 3, information technologies have become much more important across most research fields in the past two decades, but the utilization of these new tools has outpaced the ability of some fields and disciplines to develop standards and practices that will ensure a level of transparency consistent with fostering integrity and reproducibility.

How should fields go about developing new standards and ensuring that they are followed? One recent article encouraged disciplines to develop detailed case studies on selected nonreproducible publications with the goal of "deriving general principles for improving science in each field" (Alberts et al., 2015). One historical example of a field where DRPs were once widely tolerated is human language technology (HLT), which includes areas such as automated speech recognition and machine translation (Liberman, 2012). A public demonstration at Georgetown University in 1954 of a system that translated several Russian sentences into English encouraged the belief that the most significant barriers to machine translation had been overcome, yet the Georgetown system had a small vocabulary, and a limited number of grammar rules and did not represent a true scientific advance (Hutchins, 1982). After this demonstration, HLT received significant federal funding, but by the mid-1960s there was not much to show for it. The systems produced during these times could generate an impressive demonstration but performed poorly in real-world use, with output requiring extensive human post-editing. A negative evaluation of the potential of the field led federal agencies to largely end support for HLT research for almost two decades (NAS-NRC, 1966). When the Defense Advanced Research Projects Agency renewed support for HLT in the mid-1980s, a number of steps were taken to ensure that research produced clear, usable results. The results of all funded projects needed to be judged against a well-defined, objective evaluation metric, developed and applied by the neutral National Bureau of Standards (now the National Institute of Standards and Technology), on shared datasets, with the results of the evaluation revealed to the sponsor and the other investigators (Liberman, 2012). Although some HLT investigators complained at first about this "common task structure," the field quickly embraced it, strengthening its research culture as a result. The common task structure created a positive feedback loop that accelerated progress. Error rates decline by a fixed percentage every year, with advances mainly taking the form of incremental improvement. The sharing and reuse of data have become central to research practices in HLT. Advances in the field have led to products that are widely used today, such as Apple's Siri and Google Translate.

In recent years, there have been a number of positive developments related to ensuring quality and reproducibility at the broad level of the research enterprise

as well as in specific fields and disciplines (*Nature*, 2015b). Experts have made the case that integrity, quality, reproducibility, and the credibility of research are strongly interconnected:

> If science is to enhance its capacities to improve our understanding of ourselves and our world, protect the hard-earned trust and esteem in which society holds it, and preserve its role as a driver of our economy, scientists must safeguard its rigor and reliability in the face of challenges posed by a research ecosystem that is evolving in dramatic and sometimes unsettling ways. (Alberts et al., 2015)

While it would take considerable space to list or describe all the recent and ongoing efforts, it is worth identifying a few significant initiatives. A 2012 workshop identified key requirements for methodological reporting in animal studies aimed at improving the predictability and quality of preclinical animal studies, such as sample size estimation, whether the animals were randomized and how, and data handling (Landis et al., 2012). In 2013, *Nature* introduced a checklist that is "intended to prompt authors to disclose technical and statistical information in their submissions, and to encourage referees to consider aspects important for research reproducibility" (*Nature*, 2013). In biomedical research, the EQUATOR (Enhancing the QUAlity and Transparency of health Research) Network (http://www.equator-network,org) is an international initiative that promotes reporting standards aimed at ensuring transparency and reliability.

Efforts to address the issue of sharing clinical trial data have also gained momentum in recent years (IOM, 2015). For clinical trials, sharing data at the time of publication is aspirational. There may be many reasons to wait for a specified period of time before opening up the data and metadata for sharing (IOM, 2015). Several recent proposals indicate that consensus is building around a standard recommended maximum of 6 months following publication for data to be shared (IOM, 2015; Taichman et al., 2016).

The Center for Open Science, the group that was responsible for the recent effort to replicate psychology results discussed above, has also developed a set of Transparency and Openness Promotion (TOP) guidelines that it has put forward for consideration and possible adoption by journals (Nosek et al., 2015). The TOP guidelines include eight standards, with each standard comprising three levels that are intended to encourage movement toward greater transparency and openness over time. Two of the standards are intended to reward researchers for open practices by establishing citation standards for data, code, and research materials and by establishing conditions under which the journals will publish replication studies. Four of the standards specifically define openness through the research process in design standards, research materials, data sharing, and analytic methods. The final two standards cover preregistration of studies and analysis plans that are aimed at clarifying the distinction between research intended to confirm hypotheses and research intended to generate hypotheses. The TOP guidelines have already attracted an impressive list of signatory journals, including a number

of journals from outside of psychology and even general scientific journals such as *Science* and *PLOS ONE* (COS, 2015).

The examples of biomedical research, social psychology, HLT, and high-energy physics highlight the importance to research quality and integrity of reproducibility of results and the availability of data, code, and other information necessary for replication. Disciplines and fields have traditionally had a wide variety of cultures and practices related to data (NAS-NAE-IOM, 2009a). In Chapter 3, the problems caused by resistance to sharing of data and code in climate science were described. Even in some areas of computational science, where the value of transparency would appear to be obvious, there are significant barriers to reproducibility, including routine withholding of code and data on sponsored research (Liberman et al., 2012).

The efforts of the Center for Open Science and others raise the possibility that fields and disciplines can establish and implement higher standards that define today's commonly tolerated DRPs as unacceptable and provide checks and incentives to reduce the occurrence of those practices to a level far below what exists today. Progress on this front will help to foster research integrity as well as improve the quality of research across a range of fields and disciplines.

## AUTHORSHIP-RELATED CHALLENGES TO RESEARCH INTEGRITY

### Nature and Scope of the Problem

As discussed in other parts of this report, published papers are the currency of science. Through such papers, science is communicated, critiqued, and assessed. The number and quality of published articles credited to a scientist, especially peer-reviewed articles, are major criteria for promotion and tenure, and so have a powerful impact on scientific careers. Authorship designates who is willing to take responsibility for an article and who bears responsibility for the work in case of error or allegations of misconduct. Authorship credit is therefore an integral part of the scientific enterprise as a professional system.

Chapter 3 discusses how changes in the research environment such as technological advances that have transformed many aspects of performing and reporting research, the growing importance of collaborative and interdisciplinary research, and the globalization of research are affecting authorship practices and conventions. Several of the most difficult challenges to research integrity involve authorship abuses, particularly authorship credit misallocations/misappropriations (B. C. Martinson and Z. Master, personal communicaation, July 27, 2015). As discussed in Chapter 4, plagiarism is one category of authorship credit misallocation that is included in the definition of research misconduct by the U.S. federal government and by most other countries. For the most part, other categories of authorship credit misallocation are considered detrimental research practices for

the purposes of this report. This section will describe some of the most pressing challenges related to authorship and research integrity and consider the advantages and disadvantages of alternative approaches to addressing them.

Authorship can be misused in several ways. Gift, guest, or honorary authorship involves listing an author who made no substantive contribution to the research reported. For example, researchers may add the name of a prominent researcher to a paper in the belief that it will increase its odds of being accepted by a prestigious journal. Gift authorship can happen with or without the knowledge or permission of the researcher being "honored." When the gift author had no role in the conducting or writing of the article, listing his or her name is a misallocation of credit. In cases where work is fabricated or falsified, questions are raised about the responsibilities of coauthors whose contributions may or may not have merited authorship. The stem cell case at Seoul National University and the University of Pittsburgh, described in Appendix D, discusses these issues.

A senior scientist may demand or be granted an authorship designation for a "specialized" service such as providing biological materials or specimens, helping to secure funding for the research, or serving as head of the laboratory or department where the research is undertaken. Insistence by a scientist in a position of authority that he or she be listed as an author on all papers submitted to journals by subordinates, including articles in which the senior scientist has played no direct role, is known as "coercive authorship."

As data and code sharing become part of the usual practice of science, reuse of these scholarly outputs is increasingly common. The expectation is that the use or reuse of data and/or code produced by another researcher will be appropriately cited. Such recognition rewards the producer of the data and code while improving, extending, and building on these objects in their own right. It is inappropriate to condition data or code reuse on coauthorship when there is no other contribution to the paper. This is a coercive practice that slows the advancement of science when other mechanisms are in place to reward data and code contributors, such as citation. The practice of conditioning data use on coauthorship is more widespread in some disciplines than in others but should not exist in any discipline. This is separate from, and not to be confused with, a data or code contributor who is or becomes part of the research team and collects novel data or builds code for the purposes of a research project or series of projects. Coercive authorship practices occur when coauthorship is conditioned on using data and code associated with a previous or different project rather than the only expectation being citation for downstream use.

Another detrimental authorship practice is unacknowledged or "ghost" authorship, in which researchers who have made a substantial contribution to a research article are not listed as authors. Not all unacknowledged authorship fits into this category. For example, reporting someone else's research results as one's own without designating that person as an author and without their knowledge is a form of plagiarism. A professional writer whose only involve-

ment in the research is participation in writing the paper is not considered to be an author in most contexts, but many journals require that professional writers be acknowledged.

A problematic form of ghost authorship arises when researchers who are directly involved in all phases of the research are not acknowledged (Fugh-Berman, 2010). For example, a pharmaceutical company may finance and undertake research that supports a non-FDA-approved use of one of its products, prepare the paper, and recruit prominent medical researchers to sign on as authors. The corporate support and industry authors may not be disclosed. In some cases, the listed academic authors will have had some involvement with the research, but sometimes they do not. In these latter cases, ghost authorship also becomes a type of honorary authorship.

While the immediate motivation for this form of ghostwriting is to hide the financial interest of the sponsor and ghost authors in the work, it has also been associated with other detrimental research practices such as selective reporting and suppression of some findings. In the Paxil case described in Appendix D, data falsification was admitted by the sponsor and ghost authors but denied by the listed authors. If data are falsified or the reported results are misleading in such clinical studies and the listed authors are not able to vouch for the integrity of the data or results, using the study as a basis for treating patients may present serious health and safety risks.

In addition to the Paxil case, several other examples of alleged ghostwriting that involved other alleged detrimental research practices led to legal consequences for both medical industry sponsors and ghostwriters (Feeley, 2012; Fugh-Berman, 2010). In one case, documents were released showing that Pfizer's Wyeth Pharmaceutical Company had not disclosed its role in preparing journal articles supporting the used of Prempro, a hormone drug, and recruiting academic authors (Fugh-Berman, 2010). In 2012, Pfizer had paid $896 million to settle only about half of the cases alleging Prempro had caused cancer (Feeley, 2012). In addition to Paxil and Prempro, ghostwriting has "been documented in the promotion of 'Fen-phen', Neurontin, Vioxx and Zoloft" (Fugh-Berman, 2010). The companies that produce these drugs have paid millions to billions of dollars in lawsuit settlements.

This form of ghostwriting has been condemned as an "example of fraud" and "a disturbing violation of academic integrity standards, which form the basis of scientific reliability" (Bosch and Ross, 2012; Stern and Lemmens, 2011). The practice is not currently equated with plagiarism and so is not within ORI's power to regulate. Bosch and Ross (2012) suggest that ORI include ghostwriting in its definition of research misconduct so that it can be investigated and addressed under the federal research misconduct policy. The International Committee of Medical Journal Editors (ICMJE, 2015) established criteria against which to determine appropriate assignment of biomedical authorship and recommends that those who do not meet all of the criteria only be listed in the acknowledgments

sections. The Committee on Publication Ethics (COPE, 2011) also recommends that specific rules be implemented to prevent ghostwriting, which is explicitly defined as misconduct in its guidelines. The pharmaceutical industry itself has promulgated guidelines for clinical trials that specify adherence to the ICMJE authorship criteria (PhRMA, 2014).

All of the authorship abuses described above undermine research integrity. Even when the research that is reported is correct and of high quality, inaccurate and misleading authorship designations can lead to misallocation of credit, rewards, and future resources. They can damage the conduct of science if, for example, authorship credit without deep knowledge or skill in the science involved helps promote an honorary author to a position of authority. They can also obscure responsibility for reported work and make it more difficult to address other forms of misconduct, such as data fabrication. Indeed, there is evidence that engaging in authorship credit misrepresentation increases the risk that researchers will engage in research misconduct later (B.C. Martinson and Z. Master, personal communication, July 27, 2015). Several cases discussed in Appendix D, including the Paxil case and the stem cell case at Seoul National University and the University of Pittsburgh involve authorship.

Over the past several decades, surveys and meta-analyses have shed light on how prevalent inaccurate and misleading authorship designations are. A 2011 meta-analysis of research on authorship found that an average of 29 percent of respondents had experienced some problems with misuse of authorship (Marusic et al., 2011). An international survey of authors of articles published in six general medical journals in 2008 found that 21 percent of papers had honorary and/or ghost authors, down from 29 percent in 1996 (Wislar et al., 2011). Both the 2011 and 1996 surveys used the ISMJE definition of authorship (to be discussed in more detail below). Almost two-thirds of the 2011 respondents resided in the United States or Canada, with most of the rest residing in Europe. Even if other fields have a much lower incidence of authorship misrepresentation than biomedical research, the overall incidence would be disturbingly high, since biomedical research constitutes a large fraction of overall research funding and publishing.

More recent work presented at a scientific meeting and reported in the media found significantly higher rates of guest and ghost authorship than the results cited above (Jaschik, 2015).

### Addressing Authorship Credit Misrepresentation

Stakeholders in the research enterprise widely recognize that more vigorous efforts are needed to reduce and ultimately eliminate authorship credit misrepresentation. In recent years, a number of journals and professional groups such as the Council of Science Editors, COPE, and ICMJE have updated and clarified their authorship criteria to prohibit honorary and ghost authorship. Journals also

are adopting practices such as author contribution statements and are requiring independent approval of all coauthors on articles as mechanisms to discourage inaccurate authorship designation. In a 2009 report, the Institute of Medicine called on academic medical centers and teaching hospitals to prohibit medical ghostwriting (IOM, 2009).

A 2012 editorial in *Science* called for renewed attention to the problem of honorary authorship and advocated that more journals adopt the use of author contribution statements (Greenland and Fontanarosa, 2012). The editorial also called on research institutions to combat honorary authorship more directly and proactively, pointing out that institutions such as Washington University in St. Louis define honorary authorship as misconduct in their policies (Washington University, 2009). For example, junior researchers need to know who to notify and the appropriate procedures to follow when they are coerced into listing a noncontributing coauthor.

Several alternative approaches might be considered to address this challenge. One would be to treat some forms of authorship credit misrepresentation in addition to plagiarism as research misconduct. A footnote in the 1992 *Responsible Science* report states that "it is possible that some extreme cases of noncontributing authorship may be regarded as misconduct because they constitute a form of falsification" (NAS-NAE-IOM, 1992). *Responsible Science* also noted that, in 1989, a Public Health Service annual report of its activities to address research misconduct included several abuses of authorship as examples of misconduct, such as "preparation and publication of a book chapter listing co-authors who were unaware of being named as co-authors" and "engaging in inappropriate authorship practices on a publication and failure to acknowledge that data used in a grant application were developed by another scientist." It should be noted that this formulation predated the 2000 federal policy on research misconduct. In 1989, the PHS definition of research misconduct was "fabrication, falsification, plagiarism, or other serious deviations from commonly accepted research practices." None of the specific terms was further defined.

Authorship misrepresentation other than plagiarism is clearly not included in the definition of falsification specified in the current U.S. federal research misconduct policy (OSTP, 2000). A change in the definition of falsification would be needed for inaccurate or misleading authorship designations to be treated as research misconduct by the federal government.

Implementation of such a change would face a number of practical obstacles. To begin with, although the authorship standards of COPE, the Council of Science Editors, and ICMJE are widely respected, disciplines vary widely in authorship standards and practices. For example, ICMJE defines authors as those who have fulfilled the following criteria: (1) substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; (2) drafting the work or revising it critically for important intellectual content; (3) final approval of the version to be published; and (4) agreement

to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved (ICMJE, 2013a). However, in research fields involving work on complex instruments and the generation of large amounts of data, it is possible to imagine circumstances where articles are published in which no one qualifies as an author according to the ICMJE criteria. The same circumstances might imply author credit misrepresentation in one field and acceptable practice in another. This would make it difficult to develop a workable definition of falsification that could be applied in a consistent way.

Professional disputes and legal allegations over the denial of rightful authorship or a lack of rightful authorship credit have become a growing issue within the research enterprise. While academic theft is a serious transgression, it may be difficult to determine how, or from whom, an idea originated. There are numerous examples of researchers, often postdocs and junior scientists, proving that their research had been published without their name credited as an author or without their knowledge at all, both inside and outside of academia. However, there are also instances in which graduate students or junior scientists perform research with a mentor who developed the same research idea years earlier. In 1995 a graduate student, Pamela Berge, won more than $1 million in a lawsuit claiming academic theft against her mentors; however, it was later revealed that the research had been ongoing for several years before Berge entered the research laboratory and the verdict was overturned (Woolston, 2002). Clear communication and discussion of how authorship roles are to be determined at the onset of research may avoid later questions of authorship credit.

Another practical difficulty in addressing authorship credit misrepresentation other than plagiarism through the research misconduct policy framework involves the sheer scale of the phenomenon. Suppose that the study cited above is correct and more than 20 percent of biomedical research articles have honorary and/or ghost authors (Wislar et al., 2011). There are roughly 50,000 biomedical articles published by U.S. authors per year (NSB, 2012). If current practices were to continue, therefore, roughly 10,000 additional incidents of research misconduct would occur each year in just one discipline. While these incidents would certainly not all be reported or investigated, even 2,000 to 3,000 additional cases per year is more than an order of magnitude greater than the current combined number of cases now handled by NSF-OIG and ORI per year, which itself reflects substantial recent increases. By expanding the scope of the federal research misconduct definition in this way, implementing the recommendation might require significant additional resources for ORI, NSF-OIG, and perhaps other agencies.

Also, since the federal misconduct policy only applies to federally funded research, as discussed above, a change in interpretation of the research misconduct definition would not address honorary, coercive, or ghost authorship in purely privately funded research except as an exemplar and spur to raise standards across

the board. The problem of ghostwriting discussed above, for example, largely concerns research that is funded by companies.

An alternative approach to reducing and ultimately eliminating authorship credit misrepresentation would rely on identifying best practices for researchers, institutions, sponsors, and journals, and encouraging that these stakeholders accelerate adoption of these practices. For example, at the disciplinary level, societies and journals could work to update and specify their authorship standards. Sponsors and journals could more actively discourage ghost and guest authorship. A pathway toward strengthening authorship standards is discussed in Chapter 8. Chapter 9 discusses best practices, and Chapter 10 covers findings and recommendations addressing these issues.

## ADDRESSING OTHER MISCONDUCT: PROTECTING THOSE WHO RAISE CONCERNS AND MAKE GOOD-FAITH ALLEGATIONS

As discussed above, whistleblowers are a critical source of information that leads to uncovering and investigating research misconduct. Those accused of misconduct or others at their institutions often retaliate against whistleblowers—according to one survey of research misconduct whistleblowers, around 70 percent experienced some negative consequences, including more than 20 percent who lost their positions (Lubalin et al., 1995). The falsified grant application case (Appendix D) illustrates the vulnerability of whistleblowers, even in situations where there was no retaliation on the part of the accused or the institution. Providing effective protection for whistleblowers is a key element in addressing research integrity going forward (Kornfeld, 2012).

What policies and practices toward research misconduct whistleblowers are needed at the institutional level? Institutions may have policies protecting whistleblowers, although it is not clear how many actually do. Even where policies exist, it may be difficult to effectively implement protections without a strong commitment from the institution. It is often not clear (and difficult to prove) whether difficulties experienced by whistleblowers are retaliation as a direct result of making an allegation. The "tone at the top" is also very important in determining how whistleblowers are treated (Gunsalus, 1993). Chapter 9 discusses best practices in institutional policies and practices in this area, including the commitment to maintain multiple anonymous mechanisms for reporting suspicions and allegations.

Federal policies have an impact as well. As discussed in Chapter 4, under the pre-2000 federal definitions of research misconduct, retaliation against a whistleblower or other obstruction of a research misconduct investigation could be pursued by NSF-OIG or ORI under the "other serious deviation" clause. Under the current definition, the federal oversight agencies may refer allegations of whistleblower retaliation to the institution, but have no further recourse after the institution makes its report, even if they believe that there are problems. By

contrast, NSF-OIG or ORI can send back an inadequate institutional report on fabrication, falsification, and plagiarism, or in NSF-OIG's case, take over the investigation itself.

While including whistleblower retaliation as an element in the research misconduct definition is an option, there are other federal policy options that appear to be more straightforward and potentially more effective. One option would be to create standards for institutions as part of the research misconduct policy, without making whistleblower retaliation part of the misconduct definition. HHS published proposed standards for protecting research misconduct whistleblowers in November 2000 (HHS, 2000). These standards followed up on draft guidance developed by the Ryan Commission (Commission on Research Integrity, 1995). The standards were never implemented.

Another option would be to extend federal whistleblower protections to those who make allegations of research misconduct outside the federal government. This approach has actually been implemented. Research supported by the American Recovery and Reinvestment Act of 2009 (P.L. 111-5, 123 Stat. 115, 516) required recipient institutions to have whistleblower protection policies in place and specified multiple mechanisms for reporting research misconduct allegations (including to funding agency officials and members of Congress). It should be possible to look at the experience with the act and evaluate whether implementation of these protections created any difficulties for institutions, and whether this was an effective approach. Congress has the option to extend those provisions to all federal research.

In this connection, the problem of knowingly making false allegations of research misconduct deserves attention. Very little is known about the incidence of such allegations and how they are resolved. A researcher might be motivated to make a false allegation out of a desire for competitive advantage if the accused and accuser were working in the same area of research, because of commercial or political interests, personal animus, or mental illness. Bad-faith whistleblowers may have a financial incentive to make a claim; under the False Claims Act, individuals are able to sue on behalf of the U.S. government if they have "evidence of fraud against federal programs or contracts" and receive a small percentage of what is recovered (NWC, 2016).

Some personal testimony is available that provides guidance on the steps that should be taken by researchers who are falsely accused (Goldenring, 2010). Knowingly making false allegations of research misconduct is damaging in that they impair the work of the accused and his or her collaborators and also impose costs on the institutions, journals, and others who are required to investigate the allegation. In addition to protecting good-faith whistleblowers, preliminary inquiries and investigations certainly need to protect the accused; an investigation led by experts in their field should follow all claims. Like retaliation against good-faith whistleblowers, knowingly making false accusations is a form of other misconduct for the purposes of this discussion.

Even whistleblowers acting in good faith may not be very sympathetic figures, alienating colleagues and administrators. Apprehension about possible retaliation is certainly reasonable and can be expected to deter those who observe misconduct to come forward.

# 8

# Exploring New Approaches

*It is key that the public, who directly and indirectly provide the money to fund our research efforts and who are our most important ambassadors in advocating for the importance of scientific investigation, are confident of the processes we have in place. They need to know there is real value in the data that ultimately emerges.*

—C. Glenn Begley and John P. A. Ioannidis (2015)

***Synopsis:*** *This chapter reviews the benefits and costs of improved approaches to addressing research misconduct and detrimental research practices and explores several new approaches considered by the committee.*

## COSTS AND BENEFITS OF IMPROVED APPROACHES

Chapter 5 discusses the costs and consequences of research misconduct and detrimental research practices (DRPs). These costs are mainly considered from a U.S. perspective, even though it is important to remember that research misconduct and DRPs are global phenomena and that their costs are borne by researchers, institutions, funding agencies, and journals around the world. For research misconduct, costs include the direct financial impact of funding fraudulent work, the indirect costs of supporting research to extend this fraudulent work, and the cost of institutional investigations (one estimate puts this at $110 million per year) (Michalek et al., 2010). A total of several hundred million dollars a year would be a reasonable conservative estimate of the direct financial costs. In addition, there are the human costs of research careers sidetracked or ended and the reputational costs borne by institutions and collaborators. Some cases of fabrication and falsification have contributed to significant negative public health consequences and have also imposed financial and human costs. The historical case of the tobacco industry and the more recent case of vaccines illustrates that these costs may run into the millions or even billions of dollars over a period of years in particular cases.

Chapter 5 also discusses the costs that DRPs impose on the research enterprise. The financial costs of DRPs in the form of funding for research that does not produce reliable knowledge may be even larger than the costs of research misconduct. In addition, much remains to be learned about irreproducibility in research, including the extent to which DRPs are implicated and how significant

a problem it is in fields other than those where it is being actively examined, such as biomedical research and social psychology.

In contrast to the costs are the resources being devoted to preventing and otherwise addressing research misconduct and DRPs. Chapter 10 discusses education in the responsible conduct of research (RCR), including topics such as human subjects protection. One crude metric of spending by institutions on RCR education is the amount spent by institutions on subscriptions to the Collaborative Institutional Training Initiative at the University of Miami. Annual subscriptions for nonprofit organizations are $3,000, and thousands of institutions around the world subscribe.

The budgets of federal agencies that address research misconduct should also be included in this accounting. For example, the Office of Research Integrity's (ORI) budget for fiscal 2014 was $8.5 million, which includes its investigative oversight, educational activities, such as support for Research Integrity Officers (RIOs) training, and for research on research integrity (HHS, 2014b). The National Science Foundation Office of the Inspector General's (NSF-OIG) budget for fiscal year 2014 was $14.3 million, but only a fraction of this goes toward its investigative oversight and review activities. Some NSF awards in its Science and Society grants program go toward efforts related to research integrity. Other federal agencies also devote some resources to RCR education and to oversight of extramural funding.

One relatively new and significant source of support for research and other activity related to fostering research integrity is the Laura and John Arnold Foundation, which has awarded more than $80 million since 2011 to a variety of efforts, such as the Center for Open Science activities on reproducibility described above (see www.arnoldfoundation.org/grants/).

To recap, the costs of research misconduct and DRPs likely fall in a range of hundreds of millions to billions per year in monetary costs plus human and reputational costs, while investments in fostering research integrity through education, research, and development and implementation of improved standards and practices can be estimated to total tens of millions of dollars per year. Additional investments that would contribute to lowering the costs that research misconduct and DRPs impose on the research enterprise have the potential for delivering significant returns.

## THE RESEARCH INTEGRITY ADVISORY BOARD

The previous chapter reviewed several challenges facing the U.S. research enterprise as stakeholders seek to foster integrity and address research misconduct and DRPs. These include weaknesses and gaps in policies and capabilities for identifying, investigating, and addressing research misconduct; the need to develop and uphold updated research standards and practices in areas such as data sharing in response to technological advances; and the need for evidence-based

approaches to strengthening policy implementation, research environments, and incentives so that they better support responsible conduct.

For the most part, these challenges have no quick and easy solutions. They will likely require a continued, long-term effort on the part of all participants in the research enterprise.

For example, research institutions face significant challenges in ensuring that research misconduct allegations are effectively addressed and investigated. The available evidence, including presentations made to the committee, survey research, and specific published cases of deficient institutional responses, including those summarized in Appendix D, illustrate the complexity of these challenges. Addressing the unevenness in institutional policies and capabilities faces inherent obstacles. One source of difficulty is the infrequency at most institutions of cases that advance from the inquiry to the investigation stage. This means that institutional officials at a given time and place may lack hands-on experience with necessary tasks such as sequestering evidence, forming investigation committees with the appropriate expertise, orienting the committees to see the larger stakes beyond the institution and investigator, and ensuring that institutional and federal policies are followed. Effectively addressing the scientific and legal issues raised by research misconduct cases may require specialized knowledge and sensitivity. Institutional investigations can also get sidetracked when concerns about potentially bad publicity for the institution or personal relationships become considerations.

One option that this committee explored and ultimately decided to recommend is the establishment of an independent, nonprofit Research Integrity Advisory Board (RIAB) to help address longer-term challenges such as these. A similar body was recommended by the committee that authored the 1992 *Responsible Science* report, but the recommendation was never implemented (NAS-NAE-IOM, 1992). This committee considered the basic rationale for the new body, the specific tasks of the RIAB, its organization and funding, and the pros and cons of alternative courses of action.

### Functions of the Proposed RIAB That Are Not Being Performed by Existing Institutions

Federal agencies such as ORI and NSF-OIG play an essential role in addressing research misconduct and in related areas. These agencies oversee the investigative and educational activities of institutions through mechanisms that are separated from funding decisions and oversight, which is appropriate for the United States given the need for these agencies to be accountable to Congress. Federal agencies have also made valuable contributions to promoting integrity. These include the development of ORI's "The Lab: Avoiding Research Misconduct" interactive video, ORI's efforts to train RIOs, and NSF-OIG's involvement in efforts such as the Organisation for Economic Co-operation and

Development's Global Science Forum. However, working across disciplines and with various stakeholder groups to develop improved standards and practices relevant to research integrity is not part of their core missions. Also, the bifurcation between ORI's focus on biomedical research and NSF-OIG's focus on other fields constitutes a barrier to developing a unified focus on research integrity challenges. For example, much of the research on research integrity issues supported by ORI has naturally focused on conditions and topics within biomedical research, leading to a situation where relatively less is known about other fields and disciplines. Research integrity issues relevant to all federal agencies that fund or perform research need to be addressed by RIAB.

Nongovernmental organizations within the United States and around the world such as the American Association for the Advancement of Science, the Council of Graduate Schools, Association of American Universities, Committee on Publication Ethics, and the world conferences on research integrity are also making important contributions to promote responsible conduct and research integrity. Notwithstanding the value of these efforts, they tend to be ad hoc in nature or focused on specific disciplines or sectors.

The RIAB would provide a vehicle for the research enterprise, including research institutions, to address issues and challenges related to integrity on an ongoing basis across disciplines and sectors. Such issues will undoubtedly continue to arise in the global research environment and will continue to change. The committee hopes that publication of this report will stimulate ongoing dialog. Such a dialog is needed not once every few decades, as evidenced by the span between the 1992 report and this report, but on a continuing basis.

The RIAB would facilitate the exchange of information on approaches to assessing and creating environments of the highest integrity and on the handling of allegations of misconduct and investigations. For example, the effort to assess research environments that was undertaken under the auspices of the Council of Graduate Schools could be continued and expanded. Institutions might be more willing to perform such self-assessments and utilize the resulting data if their own results remain private and they are able to benchmark their environments against the aggregated results of peer institutions. The activity could also aid in the development of institutional best practices and benchmarks for the capabilities needed to respond to allegations and conduct investigations.

The RIAB would also provide advice, support, encouragement, and where helpful, advocacy on what needs to be done by research institutions, science, engineering, technology, and medical journal and book publishers, and other stakeholders in the research enterprise to promote research integrity. The body that was recommended in *Responsible Science* was expected to perform several functions, including the development of model practices, policies, and procedures for the community; the collection and analysis of data on allegations of misconduct; and the conduct of periodic studies of policies and approaches for fostering research integrity and addressing research misconduct and questionable research

practices (NAS-NAE-IOM, 1992). The RIAB recommended by this committee perform some functions that are similar to those anticipated earlier, with adjustments and details altered to incorporate the experience of the past several decades and to account for the contemporary context:

- Work with public and private research sponsors to develop improved practices and approaches to addressing research misconduct and fostering integrity. For example, the RIAB could serve as a forum for the discussion of issues where no community consensus currently exists (such as what the appropriate penalties for research misconduct should be) or where current disparate approaches should be harmonized (such as the implementation of the federal research misconduct policy in areas such as plagiarism).
- Work with science, engineering, technology, and medical journal and book publishers to develop improved practices and approaches. The biannual Journal Summit organized by the National Academy of Sciences generates a number of useful ideas that could be explored further by RIAB.
- Identify important topics and questions related to research misconduct and research integrity, including pathways to improve research environments and RCR education, where research could produce valuable insights, and perhaps serve as a mechanism for commissioning such research.
- Work with research institutions, institutional officials, and groups such as the new Association of Research Integrity Officers to identify and develop resources aimed at improving institutional capability to respond to research misconduct allegations and sustain environments that encourage responsible conduct. These resources could include just-in-time training materials, referrals to experts with relevant scientific and/or legal knowledge who could be consulted on specific cases, and help with organizing external review of investigation committee task statements and investigation reports.

As with the 1992 report, the RIAB recommended here will have no direct role in investigations, regulation, or accreditation. Rather, it will serve as a neutral resource based in the research enterprise that helps the research enterprise foster integrity in a changing environment.

### Structure, Organization, and Funding

The committee discussed several alternative organizational and funding structures for the RIAB and concluded that the RIAB would work best as an independent nonprofit organization supported by dues-paying members. While the RIAB would benefit from federal support and participation, its functions and

activities are clearly nongovernmental, so it should not be established as a federal entity. Likewise, housing the RIAB within one of the existing organizations that represent higher education institutions would impair its ability to be responsive to all relevant constituencies.

Several organizational alternatives to a stand-alone, independent RIAB were considered. One option would be for the RIAB to be affiliated with a federally funded research and development center such as the Science and Technology Policy Institute, a policy analysis organization of the Institute for Defense Analyses that works with the Office of Science and Technology Policy, the National Science Board, and other federal entities. However, since the RIAB's activities would not focus primarily on policy analysis, and affiliation with a federally funded research and development center would imply reliance on federal funding, such an arrangement would not be a good fit.

Another possibility explored by the committee would be for the RIAB to be affiliated with, or its functions performed by, the proposed Research Policy Board (RPB). In 2015 a committee of the National Academies of Sciences, Engineering, and Medicine recommended that the RPB be created to serve as an "analytical, anticipatory, and coordinating forum on regulatory matters" affecting research universities (National Academies of Sciences, Engineering, and Medicine, 2016). The proposed RPB would be situated outside government yet involve both research universities and federal agencies in its activities. However, the committee does not believe that having the RPB take on the tasks and functions of the RIAB would be an acceptable alternative. The RPB would have a broad mandate regarding regulatory issues affecting research universities, making it difficult to maintain a consistent focus on research integrity issues. The RPB would also have a closer relationship with federal agencies than would be desirable for the RIAB.

The committee believes that the RIAB could function effectively with a small permanent staff of three or four people, supplemented by fellows and consultants. An annual budget of about $3 million would be adequate. The RIAB would be governed by its members, with a rotating executive committee selected to develop strategy and oversee operations. Funding would come in the form of regular contributions from members such as the major public and private sponsors of research (NSF, NIH, Department of Defense, Department of Energy, Howard Hughes Medical Institute, and others), universities and other research institutions, industrial members, scientific societies, and science, engineering, technology, and medical journal and book publishers. Contribution amounts could be set at different levels depending on organizational types or sizes. For example, the contributions from research institutions could be set as a percentage of their annual research activity.

The ability of the RIAB to attract member participation and support would be an encouraging indicator that the community sees it as a mechanism that can help the community address the problems and issues identified in this report.

### Consideration of Alternatives

The committee also considered alternatives to setting up the RIAB or a similar body. One alternative would be to maintain the status quo, which has been characterized by reactive, ad hoc responses to high-profile cases. The status quo approach would guarantee that the research enterprise and its components do not have a mechanism that facilitates improved performance over time. It also means that there would be no organizational focus on lowering the risk that future research misconduct cases and detrimental research practices will cause serious damage to the enterprise in terms of lost credibility, wasted resources, harms to research subjects, and a slower advance of knowledge.

Another alternative would be for federal agencies to develop a more extensive framework of regulations covering institutions in the area of research integrity. For example, institutions could be required to form committees on research integrity similar to the institutional review boards that oversee experiments on human subjects. Certification and related training could be required for those holding specific positions of responsibility in institutions, such as research integrity officers. Institutions could be required to submit detailed annual reports on their RCR educational programs and related efforts. Research funding could be tied to compliance.

The committee rejected an approach relying on increased regulatory oversight of institutions for several reasons. First, during the course of the study, the committee observed that the framing and discussion of research integrity issues still tend to be focused on reacting to individual, high-profile cases of misconduct. The committee believes that more focus and effort should be devoted to encouraging integrity, rather than just reacting to misconduct, than is the case today. Adopting an enhanced regulatory approach would represent a move in the opposite direction. Second, the costs of enhanced regulatory approaches in terms of resources and institutional staff and faculty time are significant and certain, while the benefits are more difficult to estimate and cannot be guaranteed. It also is not clear that such an approach would improve institutional performance across the board. Third, institutions already carry heavy regulatory burdens related to research (National Academies of Sciences, Engineering, and Medicine, 2016). Adding to these burdens could have the effect of undermining scientific productivity throughout the system.

### Considering International Examples

International examples exist that might inform the establishment of the RIAB and its functions. For example, several countries have established research integrity organizations that are independent of government or represent collaboration among the nation's primary funding agency, the national science academy, and the national association of research universities. In addition to promoting research integrity, some of these bodies have an investigatory or adjudicatory function.

For example, the National Board for Research Integrity in the Netherlands hears appeals of the findings of institutional investigations. The RIAB recommended here (and in 1992) would not perform any investigatory or adjudicatory functions.

Another international example comes from a 2010 report of the Council of Canadian Academies recommending that a Canadian Council for Research Integrity be established (CCA, 2010). The Council for Research Integrity would have been charged with providing confidential advice, gathering and reporting information, and developing best practices in education and assessment. In 2011, Canada adopted a Tri-Agency Framework: Responsible Conduct of Research and has established the Secretariat on Responsible Conduct of Research, which performs some of these functions, but also carries other responsibilities that would not be taken up by the RIAB related to human subjects protections and to reviewing institutional investigation reports on behalf of Canada's research funding councils (SRCR, 2016).

The United Kingdom Research Integrity Office (UKRIO) is an independent advisory organization that has a structure and purpose that are relevant to the RIAB (UKRIO, 2017). Launched in 2006, UKRIO is an independent nonprofit that aims to "(1) Promote the good governance, management and conduct of academic, scientific and medical research, (2) Share good practice on how to address poor practice, misconduct and unethical behavior, and (3) Give confidential, independent and expert advice on specific research projects, cases, problems and issues."

The committee is not aware that any formal evaluation of UKRIO and its effectiveness has been done. It does much of its work in confidential, behind-the-scenes settings. In the aftermath of prominent UK research integrity cases, particularly the measles, mumps, rubella virus–autism case (see Appendix D), there have been calls for stronger institutional responses to research misconduct in the United Kingdom (Dyer, 2011). It is important to note that the UK's research funding bodies, the research councils, have no organizational equivalents to ORI or NSF-OIG that can require institutions to respond to misconduct allegations relating to publicly funded research.

## STRENGTHENING AUTHORSHIP STANDARDS AND PRACTICES

What additional steps should stakeholders in the research enterprise take to address the challenges discussed in Chapter 7? For example, how should detrimental research practices related to authorship, such as coercive authorship, gift authorship, and unacknowledged ghost authorship, be discouraged and reduced? These practices impair the usefulness and reliability of authorship as the central institution for assigning credit for reported work, fixing responsibility for that work's quality and integrity, and communicating critical information that allows other researchers to replicate, extend, and where necessary, correct that work.

The status quo is increasingly problematic. Although some disciplines have

developed clear guidelines, authorship practices and conventions are largely left to individual institutions and journals. Greater clarity at the disciplinary level about the roles that merit authorship, the contributions that do not merit authorship, the significance of author order, and the responsibilities of a primary or corresponding author would be very helpful in facilitating appropriate decisions and practices in labs and collaborations. Universal condemnation (i.e., by all disciplines) of gift or honorary authorship, coercive authorship, and ghost authorship would also contribute to changing the culture of research environments where these practices are still accepted. Universal adoption of the requirement that all authorship roles be disclosed, as is the case for a growing number of journals, and commitment to the principle that all contributors who merit authorship should be listed would also be positive steps.

## A Framework for Disciplinary and Interdisciplinary Authorship Standards

As discussed above, a number of scientific societies, journals, associations, and research institutions have developed or updated their authorship criteria and guidelines in recent years. Some of these criteria and guidelines explicitly call for an end to such practices as ghost and gift authorship.

We have good examples of authorship guidelines and standards set at the field or disciplinary level or by individual journals. Standards may also describe the responsibilities of authors in areas such as data sharing, as well as the roles and responsibilities of reviewers. For example, the journal *Neurology* has a very detailed set of authorship guidelines on its "Information for Authors" page (*Neurology*, 2016). The World Association of Medical Editors and the International Committee of Medical Journal Editors also have developed authorship standards. As explained above, these standards have some important differences. The committee favors an approach that authorship should be established through a significant intellectual contribution to the work in at least one area, such as planning, performing, analyzing, or writing. All authors should have the opportunity to approve the final manuscript.

The committee recognizes that flexibility in the development and implementation of authorship guidelines is needed due to the significant differences between disciplines. For example, in many disciplines, research is performed in complex collaborations of large, distributed groups that perform highly specialized tasks. The recent article reporting on the first observation of gravitational waves, which had been hypothesized by Einstein, is a good example of such work (Abbott et al., 2016). The article has around 1,000 authors. In such efforts, researchers who perform critical tasks in conceptualizing the work or parts of the work may not participate in collecting or analyzing data. Likewise, it is impractical for hundreds or thousands of coauthors to play meaningful roles in writing or editing a journal article. Disciplines need to be able to define for themselves what a significant intellectual contribution is. Also, manuscript approval and

specification of author roles may need to be implemented by groups or subgroups of authors through a defined procedure rather than by individuals.

The process of developing and promulgating authorship guidelines may differ by discipline. For example, the Guidelines for Responsible Conduct Regarding Scientific Communication adopted by the Society for Neuroscience (SfN, 2010) are very detailed and cover a range of issues. An SfN working group developed the guidelines, which were approved by the SfN Council. Other fields and disciplines might also develop standards through a leading society, through a coalition of societies and journals, or through another process aimed at ensuring broad buy-in by the community. Standards could also be developed in interdisciplinary areas where there is enough research activity and enough disparity in practices between the collaborating disciplines to warrant such a step. In developing interdisciplinary standards, scientific societies, interdisciplinary journals (e.g., *Science, Nature, PNAS, PLOS*), and sponsors of interdisciplinary research can play important roles.

Research institutions can make an important contribution to stronger authorship standards. A number of institutions already have adopted guidelines that prohibit practices such as guest or honorary authorship, with Harvard Medical School (1999) being a good example.

The committee believes that the widespread development and dissemination of such standards will make a significant contribution to research integrity, and urges the research enterprise to continue and accelerate progress.

The following framework for developing authorship standards outlines several baseline requirements and might be useful to disciplines that are developing or updating their standards. At the same time, the framework is flexible enough to accommodate the significant differences that exist between disciplines in their authorship practices. The committee recommends that disciplines adopt standards compatible with this framework.

*Standards should specify the appropriate roles that merit designation as an author:*

- Substantial intellectual contribution to conceiving, designing, or planning the research to be reported;
- Substantial intellectual contribution to acquiring, analyzing, or interpreting the primary data;
- Substantial intellectual contribution to drafting or revising the article reporting the research in question.

*Standards should specify the contributions that do not merit authorship but may merit acknowledgment and/or citation:*

- Securing funding for the research;
- Providing general supervisory or administrative support for the research;

- Technical writing, editing, and proofreading of the article reporting the research;
- Making available data collected for previously reported work, or providing materials or specimens.

*Standards should explicitly identify detrimental authorship practices that are unacceptable:*

- Gift or honorary authorship, coercive authorship, and ghost authorship.

*Standards should also specify:*

- That all authors should approve the final manuscript;
- That one or more authors who are accountable for the entire work should be identified;
- That the roles of each of the listed authors should be specified, including which authors or groups of authors are responsible for which aspects of the reported work;
- The types of work being covered by the standards (e.g., only primary research articles or other types of work as well);
- The process for gaining approval of articles for publication and the principle underlying the approval process (e.g., all listed authors must individually approve a manuscript prior to submission, or an alternative approval mechanism for large collaborations);
- The meaning of "substantial intellectual contributions" to relevant research in that discipline;
- The significance (if any) of author order.

## Alternative Approaches

The committee considered several alternatives to its recommended approach. One possible alternative would be for this committee or another body to develop and implement a more detailed uniform authorship standard across all disciplines. As covered above in the discussion considering whether forms of authorship misrepresentation other than plagiarism should be included in a revised federal research misconduct definition, developing a uniform authorship standard that would be meaningful and at the same time applicable to current conditions in all fields and disciplines would be impractical and probably counterproductive.

Another alternative would be to move away from the concept of authorship entirely toward a new principle for assigning credit and responsibility for reported research. The institution of authorship within research emerged with the first scientific journals in 17th-century Europe. From that time until fairly recently, the predominant mode of research production was for an individual investigator to report on experiments or observations, perhaps assisted by students, in a laboratory or field setting. As discussed in Chapter 3, some research still fits this

TABLE 8-1 Project CRediT Terms

| Contributor Role | Role Definition |
|---|---|
| Conceptualization | Ideas; formulation or evolution of overarching research goals and aims. |
| Methodology | Development or design of methodology; creation of models. |
| Software | Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components. |
| Validation | Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs. |
| Formal analysis | Application of statistical, mathematical, computational, or other formal techniques to analyze or synthesize study data. |
| Investigation | Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection. |
| Resources | Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools. |
| Data curation | Management activities to annotate (produce metadata), scrub data, and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later reuse. |
| Writing—original draft preparation | Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation). |
| Writing—review and editing | Preparation, creation, and/or presentation of the published work by those from the original research group, specifically critical review, commentary, or revision—including pre- or postpublication stages. |
| Visualization | Preparation, creation, and/or presentation of the published work, specifically visualization/data presentation. |
| Supervision | Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team. |
| Project administration | Management and coordination responsibility for the research activity planning and execution. |
| Funding acquisition | Acquisition of the financial support for the project leading to this publication. |

Source: Brand et al., 2015.

traditional paradigm, but a growing proportion of scientific activity does not. The past several decades have seen a notable shift toward larger teams, collaborations between groups dispersed throughout the world, and increased specialization. The challenges that the research enterprise faces in the area of authorship are exacerbated by a tension between the conventions of authorship, which assume a unitary authority who can vouch for the entirety of the work, will receive most of the credit for it, and decide who else will be recognized and how, and the way a significant fraction of research activity is actually undertaken today.

Some experts have advocated that the institution of authorship be replaced by a new concept known as *contributorship*. In a contributorship framework, all the contributions to reported work are identified within an agreed taxonomy. An example of work in this area is the Contributor Roles Taxonomy project (Project CRediT), which was launched following a 2010 workshop at Harvard University and aims to "provide transparency to the contributions of researchers to scholarly published work, to enable discoverability and to improve attribution, credit, and accountability" (CASRAI, 2016). Table 8-1 shows an early version of the contributorship taxonomy (Brand et al., 2015). It specifies a number of roles that are included in traditional definitions of authorship, such as conceptualization and validation, and also lists contributions such as securing funding and general supervision that are not considered appropriate author roles but might be included in acknowledgments today. The taxonomy is now being tested; Cell Press is encouraging authors to describe their contributions in this way.

An additional advantage of contributorship is that it would take advantage of the emerging digital infrastructure that automates recognition and verification systems through mechanisms like unique author (or contributor) identifiers. These systems have the potential to connect researchers with their research products so that datasets and other nonarticle contributions can be more easily utilized, and researchers who make these contributions can receive credit (CASRAI, 2016). Since the contributorship approach is inherently more transparent and less hierarchical than traditional authorship, moving in this direction might ameliorate some of the problems that have been identified in recent years related to misuse of bibliometric indicators such as the Journal Impact Factor (Alberts, 2013).

The committee believes that Project CRediT and other efforts to develop new models to modify or replace authorship are worthwhile and have the potential to make a significant contribution. They can help improve the transparency and accuracy of how credit and accountability for scientific work are assigned and recognized. The committee decided not to recommend that the research enterprise adopt the contributorship concept at this time, due to concern about including traditional author roles and other contributions within a single framework. Some committee members also strongly believe that a credit/responsibility framework for science needs to identify one or more individuals who are accountable for the entire work.

# Part Three

# Fostering Integrity in Research

# 9

# Identifying and Promoting Best Practices for Research Integrity

*An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.*

—Jonathan Buckheit and David Donoho (1995), paraphrasing Jon Claerbout

The promotion of responsible research practices is one of the primary responses to concerns about research integrity. Other responses include the development of policies and procedures to respond to allegations of misconduct (covered in Chapter 7) and education in the responsible conduct of research (covered in Chapter 10). Exploring best practices in research helps to clarify that promoting these practices is not only a moral imperative but is also essential to good science.

Over the past three decades, government agencies, advisory bodies, scientific societies, and others have issued reports, educational guides, and other materials that address the topic of research practices. For example, the 1992 report *Responsible Science* points to a number of factors that affect research practices, including general scientific norms, the nature and traditions of disciplines, the example of individuals who either hold positions of authority or command respect, institutional and funding agency policies, and the expectations of peers and the larger society (NAS-NAE-IOM, 1992). That committee's review of research practices focused on four areas: data handling (including acquisition, management, and storage); communication and publication; correction of errors; and research training and mentorship. The report explained how commonly understood practices in each of these areas promote research integrity.

A number of other documents and codes of conduct from around the world have specified good or appropriate research practices (CCA, 2010; DCSD, 2009; ESF-ALLEA, 2011; ICB, 2010; IOM-NRC, 2002; MPG, 2009; NHMRC-ARC-*Singapore Statement*, 2010; TENK, 2002; UA, 2007; UKRIO, 2009). In addition, responsible research practices have constituted the primary subject matter for responsible conduct of research education activities, as illustrated by various educational guides (Gustafsson et al., 2006; Steneck, 2007; NAS-NAE-IOM, 2009b; IAP, 2016). These materials address the topics covered in *Responsible Science*—

data handling, publication, correcting errors, and mentoring. Some add other topics, including research collaboration, peer review, conflicts of interest, and communicating with the public. Formulations of responsible research practices specific to certain fields address additional requirements, such as protection of human research subjects, care of laboratory animals, and prevention of the misuse of research and technology. For example, the National Institutes of Health (NIH, 2009) has specified nine core areas of responsible conduct of research instruction.

Given the extensive effort to formulate responsible research practices, what does this report hope to add to the discussion? One goal is to reexamine the primary elements of responsible research practices in light of current conditions for doing scientific and scholarly work. A key conclusion of this study is that significant threats to research integrity exist in the United States and elsewhere, arising from a combination of factors present in the modern research environment. As discussed elsewhere, determining the incidence and trends of research misconduct and detrimental research practices is difficult or impossible with the existing data. However, failure to respond effectively, or in some cases an apparent tolerance for detrimental research practices by researchers, research institutions, journals, and funding agencies, has clearly contributed to delays in uncovering misconduct in several well-publicized cases. In some instances, this misconduct occurred over many years, and fabricated results were reported in many papers. And while survey data have limitations, a growing number of studies indicate that the prevalence of detrimental and questionable practices is too high and that the adherence to responsible practices is too low, both in general and in particular fields that are facing problems with irreproducibility of reported results (John et al., 2012).

One reason that holding to best practices is such a challenge and is ultimately so important is that researchers, research institutions, journals, and sponsors may face incentive structures that are not completely aligned with the responsible practice of research. While individual researchers have long been recognized and discussed as potentially conflicted, it is reasonable to apply this perspective as well to other actors. For example, externally funded research is a revenue stream for research institutions and plays a business function in those settings, in addition to providing the necessary funding for scientists to conduct research. The need for institutions to maximize such funding streams may sometimes detract from their ability to uphold best practices. Institutions may not exercise the necessary degree of skepticism and oversight toward researchers who are very successful and valuable to the institution in terms of securing resources or enhancing its reputation.

Likewise, journal publishers and the editors who work for them may have incentives to take actions that are not consistent with best practices for fostering research integrity. In particular, the rise of bibliometric indicators such as the journal impact factor may pose difficulties as journal editors seek to publish the best research but also have an incentive to see the impact factor of their journals

rise as far as possible. The inappropriate practice known as coercive citation, in which authors are pressured by journals to cite other papers from the journal, is an example (Wilhite and Fong, 2012).

Finally, sponsors of research and users of research may be subject to pressures or incentives of their own that are not completely aligned with maintaining the integrity of science.

One element of this committee's task was to address the question of whether the research enterprise itself is capable of defining and strengthening basic standards for scientists and their institutions. A critical aspect of this question is that the integrity of the research enterprise is achieved not solely through the integrity of individual researchers and their research practices but through the integrity of the system of which they are a part—the combination of participants and processes that constitute the system as illustrated in Figure 1-1. The best practices outlined here aim to reflect best practices in the context of the entire system of research and the interdependence of researchers, research institutions, funding agencies, journals, societies, and other participants. Developing this updated framework of responsible research practices will help the research enterprise identify particular practices that should be better understood and adhered to and how such understanding and adherence might be promoted and fostered.

## FRAMING BEST PRACTICES FOR RESEARCH INTEGRITY

As described in Chapter 2, the values of objectivity, honesty, openness, accountability, fairness, and stewardship underlie the effective functioning of research. These values are realized through the norms that apply to research practices. For example, honesty requires that researchers do not alter the data an experiment has produced, and openness means that researchers share the methods they used.

Norms permeate research. Some are formal and explicit, such as the regulatory requirements for treatment of animal and human subjects. Others are informal and sometimes implicit. For example, although there may be no policy that explicitly prohibits practices such as taking undeserved credit for the work of graduate students or postdocs that one is supervising or not extending deserved credit to them, researchers who exploit those who they supervise for personal ends are working against the norms of science.

Norms can be descriptive as well as aspirational. Descriptive norms are those that are generally adhered to and are expected of members of the enterprise. Sanctions may be attached to serious violations of descriptive norms; for example, all those involved expect that researchers will accurately report the results of their research. Aspirational norms are ideals that members of the research enterprise hold and attempt to achieve; for example, researchers seek excellence in the design and execution of their research and seek results that will make significant contributions to the body of knowledge in a field (Anderson et al., 2010).

The best practices described here are aimed at individuals and entities serving different roles within the research system, including researchers, reviewers, institutions, journals, and funders. The committee uses the term *best practices* here to refer to prescriptive and aspirational norms. The committee has drawn these best practices from the relevant literature, from the experts that it has consulted, and from the accumulated knowledge and experiences of its members. The practices identified encompass principles, strategies, modes of behavior, and activities that preserve the integrity of research and avoid the pitfalls that impede scientific progress. Except where noted, these practices do not require significant additional resources to implement and are indeed practiced in a variety of locations and settings. For most of these practices, the necessary conditions for implementation are recognition on the part of the identified stakeholders that the integrity of research is central to the practice and progress of research, and willingness to act on that recognition. One of the major impediments to such recognition and willingness, of course, is that these practices may not be completely aligned with the perceived self-interests of some stakeholders.

These best practices do not cover every possible ethical situation encountered in research. Nor do they include matters of science and technology policy that are largely administrative, procedural, or discipline specific, such as data retention policies in particular fields or the distribution of research funds. However, the ethical and the administrative overlap in many areas, especially in areas involving obligations of stewardship to the research system as a whole (e.g., in workforce policies), and these overlapping areas are addressed in what follows.

These best practices apply across all areas and forms of research. In contrast, specific codes of conduct are more prescriptive than best practices and can vary from discipline to discipline, such as the number and order of authors on a paper. The application of best practices may also vary in some particulars depending on whether research is undertaken in academia, industry, or government laboratories. The following compilation will strike many readers who are experienced in research as self-evident. These responsibilities are delineated here in part to demonstrate the dense web of relationships and obligations that characterize the research enterprise.

The committee has aimed to describe best practices that are specific enough to be implemented but that may also encompass a number of detailed components. Responsible research practice checklists are provided to enumerate these components.

## Researchers

Principal investigators and other scientists (including technicians, undergraduate and graduate students, and postdocs) are the foundation of the research enterprise. The research record begins with their work, and researchers are the primary evaluators and verifiers of work done by others in their respective

fields. Every scientific finding a researcher reports contributes to progress in the discipline, and failings made in the conduct or reporting of the research can immensely harm the progress of the field. Every researcher has the responsibility to ensure that these tasks are carried out to the best of his or her ability.

Researchers may play a number of roles during their careers, often simultaneously, including student, trainee, young investigator, principal investigator, department head, reviewer, editor, and administrator. The research process itself includes planning research, performing research, and disseminating results, and researchers have responsibilities at all points during the process. In planning research, they need to consider the effects of research, both positive and negative, on the broader society. It is especially important that they be vigilant about the possibility of unanticipated and potentially dangerous consequences of research, whether on a local or global scale. In interdisciplinary or international research collaborations, investigators may need to engage in continuing discussions about the standards that apply to such efforts.

As they perform research, scientists are expected to maintain high standards of proof and scientific credibility through validation of methods and rigorous confirmation of findings. They should keep clear and accurate records. They should follow the rules and procedures of their institution and laboratory regarding the physical and electronic security of data and the devices on which they are stored. They need to adhere to policies and regulations on the conduct of research related to personal safety. They should be open with supervisors and funders regarding progress, including positive and negative results.

Disseminating research entails responsibilities as well. Researchers should give credit to colleagues for help in completing work, whether in a presentation or a manuscript. They should reveal all methods and corresponding experimental findings that support conclusions as well as any unexplained outlying data that do not fit with the conclusions, allowing others to decide whether the conclusions are still valid despite the outliers.

**Best Practice R-1: Research Integrity. Uphold research integrity with vigilance, professionalism, and collegiality.**

According to one formulation, integrity for the researcher "embodies above all the individual's commitment to intellectual honesty and personal responsibility" (IOM-NRC, 2002). The duty of researchers to uphold research integrity is multifaceted. Fulfilling this duty starts with a broad understanding of scientific methods and the research enterprise as a human institution. Research requires the constant exercise of judgment and is subject to bias, whether conscious or unconscious. Researchers need to be aware of their own personal potential sources of bias in designing, carrying out, evaluating, and reporting their own work. They need to understand that knowledge advances over time, although errors and mistaken interpretations can occur along the way. Researchers who acknowledge and correct their own errors or misinterpretations with equanimity contribute to

the progress of science. Likewise, researchers should be fair and generous when critiquing the work of others. Criticisms should focus on errors in the work and disagreements about interpretation, but not on the person.

In addition to meeting their field's standards of integrity and quality in their own work, as specified in the best practices on data handling and authorship, researchers need to promote high standards among colleagues. They should take careful and timely action when a concern about research integrity arises. As a prerequisite, they should understand the definitions of, and policies to address, research misconduct adopted by their institutions and funding agencies. They should be familiar with the appropriate formal procedures for expressing concerns and making allegations, as well as informal rules and steps to help ensure that such concerns and allegations are made responsibly (Gunsalus, 1998a). These informal rules include accounting for one's own biases, appreciating that one's knowledge of a situation may be incomplete or incorrect, and getting confidential perspectives on possible misconduct from a trusted advisor before making a formal allegation.

Researchers should maintain an active commitment to openness in research as the essential foundation of academic freedom, not just the integrity and credibility of science. A commitment to openness means both acting and advocating for openness.

**Best Practice R-2: Data Handling. Manage research data effectively, responsibly, and transparently throughout the research process. This includes providing free and open access to research data, models, and code underlying reported results to the extent possible, consistent with disciplinary standards, funder requirements, employer policies, and relevant laws and regulations (such as those governing intellectual property).**

Effective record keeping and data management while undertaking research, and complete sharing of data, models, and code when publicly reporting results, are fundamental to research integrity. The importance of updating knowledge and practices related to data is increasingly recognized around the world (NAS-NAE-IOM, 2009a; KNAW, 2013). The pitfalls that can occur when dishonest, closed, or ineffective data management practices are employed are illustrated by the translational omics case and other examples discussed in Chapter 7 and Appendix D.

Researchers need to understand and follow the data collection and analysis standards of their own fields. For example, research data will often contain potential outlying results. While refining data to remove outliers is appropriate, any data refinements should be made to the entire dataset and should similarly improve subdatasets as it does the entire set. The refinement should also be well documented wherever the dataset appears. Some data refinements made after an experiment may be acceptable, since the types of noise that will show up in a dataset may be unclear until after the data are collected, but should be based on an analytic principle that provides an explicit rationale for exclusion. Researchers should guard against the temptation to use a post hoc rationale to make undocu-

mented refinements that strengthen support for a favored hypothesis. Such behavior is a detrimental practice or could even cross the line and become falsification.

In some settings and some cases, data, models, and code may not be made available, or sharing may be delayed due to legal or regulatory restrictions, including those related to privacy, intellectual property protection, and national security classification. For research that does not result in publicly reported results, such as some work performed by industrial or government labs, sharing of data and code is not a requirement but should be undertaken where possible.

In the 21st century, many novel findings and published works are based on nonobvious analysis of large datasets. How to effectively manage these datasets and properly provide them or refer to them during review and publication are challenging issues that are being considered across many fields and disciplines. Internal curation of large datasets may be expensive for research groups, and many journals do not have resources to host the datasets. However, examples of falsification, fabrication, or error discussed in Chapter 7 illustrate that posting of data and code can enable researchers to identify problematic conclusions and correct the research record.

Researchers need to ensure that appropriate statistical and analytical expertise is utilized in the project. The use and misuse of statistical tests such as *p*-values are current topics of discussion in a number of fields; the American Statistical Association recently released a statement listing six principles on the misconceptions and misuse of the *p*-value (Wasserstein and Lazar, 2016). Researchers should avoid detrimental practices such as *p*-hacking, in which statistical and analytical parameters are adjusted until a desired result is achieved (Nuzzo, 2014). Supervisors should stay close to the primary data even if they lack the technical skills to generate those data themselves.

**Best Practice R-3. Authorship and Communication. Follow general and disciplinary authorship standards when communicating through formal publications. Describe the roles and contributions of all authors. Be transparent when communicating with researchers from other disciplines, policy makers, and the broader public.**

Decisions about authorship of research publications are an important aspect of the responsible conduct of research. Although many individuals other than those who conceive of and implement a research project typically contribute to the production of successful research, authors are considered to be the person or persons who made a significant and substantial intellectual contribution to the production and presentation of the new knowledge being published.[1]

[1] In Recommendation Five, this report calls for the development and adoption of authorship standards and suggests a framework that if adopted would formally codify several of the best practices discussed here, such as describing the roles of all authors. See Chapter 8 for the rationale underlying the recommendation and Chapter 11 for the recommendation text.

As discussed in Chapter 3 and Chapter 7, authorship is also the "coin of the realm" in science—the mechanism through which scientists receive credit for intellectual work. Authorship, particularly lead authorship, carries with it credit that affects careers and promotions. Because of this, authorship often becomes a fraught topic and can invite misconduct and detrimental research practices.

In addition, authorship carries responsibilities. For example, authors are responsible for the veracity and reliability of the reported results, for ensuring that the research was performed according to relevant laws and regulations, for interacting with journal editors and staff during the publication, and for defending the work following publication (Smith and Williams-Jones, 2012). The article or paper presented by researchers "should be complete, and, where applicable, include negative findings and results contrary to their hypotheses" (NHMRC-ARC-UA, 2007). Publication bias, selective reporting, and poor reporting are serious problems that damage the research record. Authors also need to follow discipline-specific reporting guidelines, such as those covering the registration and reporting of clinical trial results. They are responsible for ensuring that previous work is appropriately and accurately cited. In all fields, responsible authorship involves avoiding detrimental practices such as honorary authorship and duplicate publication, as well as the affirmative responsibility to ensure that all who deserve credit on a paper receive it.

As discussed in Chapter 3, authorship practices vary among disciplines and within research groups and may change over time; professional and journal standards and policies on authorship also vary (journal best practices are discussed below). Technological changes in how research is done and the prevalence of multidisciplinary and even global research teams have raised challenges for authors, such as an increase in the number of authors per paper and more limited knowledge by all authors of the methods used by other contributors.

Authors should clearly identify which portion of a research project each co-author performed (see the section on best practices for journals below). Even in cases where this is not required, this information can help readers interpret the work and may also avoid blanket condemnations if the work is later shown to be flawed. If responsibility for an article or other communication is not specified as clearly as possible, all authors can be held accountable for its contents.

Researchers may also need to communicate with specialists from other fields in interdisciplinary studies or may have opportunities to explain their work to policy makers and the broader public. Similar standards of accuracy and transparency should apply. For example, "any attempt to exaggerate the importance and practical applicability of the findings should be resisted" (ESF-ALLEA, 2011). The authors of a research article or other communication have a responsibility to ensure that press releases and other institutional documents describing that work are accurate and unexaggerated. Researchers should work with their institutional media affairs office to avoid unfounded claims and reveal both the positive and the negative aspects of research results. Researchers should also become more

sophisticated in distinguishing between reporting research results and advocating policy positions related to their research. Issues of advocacy can be complex, and no hard-and-fast rules cover all situations.

**Best Practice R-4: Mentoring and Supervision. Know your responsibilities as a mentor and supervisor. Be a helpful, effective mentor and supervisor to early-career researchers.**

The 1992 report *Responsible Science* defines a mentor as "that person directly responsible for the professional development of a research trainee" (NAS-NAE-IOM, 1992). In this report, the term *supervisor* is used to describe the person directly responsible for the professional development of a trainee. Here, the term *mentor* refers to a broader group that includes supervisors as well as other more senior researchers who are in a position to contribute to the professional development of trainees and junior researchers. Professional development encompasses the development of technical expertise, socialization in research practices, and adherence to the highest standards of research integrity. The 2002 report *Integrity in Scientific Research: Creating an Environment That Promotes Responsible Conduct* outlines the responsibilities of supervisors as including "a commitment to continuous education and guidance of trainees, appropriate delegation of responsibility, regular review and constructive appraisal of trainees, fair attribution of accomplishment and authorship, and career guidance, as well as help in creating opportunities for employment and funding" (IOM-NRC, 2002).

Since supervisor-trainee relationships are often complex, it is important that supervisors and trainees clarify their mutual expectations for the relationship (NAS-NAE-IOM, 2009b). Conflicts can sometimes occur over the time and opportunities allocated to trainees, credit for and ownership of results, and other issues related to research practices. Supervisors should make sure that trainees are aware of the risks of misrepresenting data, should be aware that subordinates can have an overzealous concern to meet expectations, and should recognize that periods of heightened stress may impair their judgment.

In the context of this report, ensuring that trainees understand and follow best practices in research is an important element of mentorship. This includes checking the work of trainees, particularly work that is being submitted for publication. In several of the individual cases that the committee examined during the study, failures and deficiencies in mentorship and supervision were factors contributing to significant delays in addressing serious problems with data underlying reported results.

Supervisors and other mentors should ensure that trainees receive high-quality instruction in, and appropriate socialization into, the responsible conduct of research. This may involve incorporating activities within the lab as well as institutional and other instruction. A potentially useful practice is to set aside portions of group meetings to discuss issues of research integrity, including group analysis of current examples of detrimental practices. Supervisors should

be certain that all persons working under them understand their commitment to responsible research and their expectation for responsible conduct. Students, researchers, and staff should be encouraged to be open about results. Constructive skepticism serves a valuable function in research. "Show me the data" is always a legitimate request. Supervisors should cultivate the expectation that others in the group may be asked to confirm complex experiments or unexpected findings, not as a check on the individual competence or integrity of research group members, but as needed to ensure validity.

In addition to the formal supervisory relationships discussed above, mentoring occurs informally in many cases. Individuals may have multiple mentors, both formal and informal, and all have some responsibility for the appropriate socialization of those they mentor. Mentors should be sensitive to the challenges that mentees belonging to underrepresented groups may be facing. Mentors need to avoid the reality and even the appearance of exploitative practices, such as asking graduate students to babysit or house sit. Although the responsibility of avoiding hypercompetitive research environments characterized by intense resource competition lies mainly with institutions and sponsors, as described below, individual supervisors should do what they can to prevent competitiveness in the lab from reaching the point where it becomes harmful.

**Best Practice R-5: Peer Review. Strive to be a fair and effective peer reviewer who provides careful reviews, maintains confidentiality, and recognizes and discloses conflicts of interest.**

Peer reviewers of grants and journal submissions provide the guiding and corrective machinery that enables the research enterprise to progress. As in other contexts of their work, researchers who serve as reviewers are expected to be honest, objective, and accountable and to preserve confidentiality and protect the ideas of others during the review process. In the context of grant review, peer reviewers are responsible for determining whether a research direction is worthy of funding based on novelty, importance, available data, and whether the proposed methods are suitable for the investigation. For journal submissions, the reviewer's responsibility is to carefully evaluate the experimental design, presented data, and analysis techniques to determine whether they cumulatively support the presented interpretation and conclusions from the data.

Potential reviewers should completely disclose conflicts of interest to the program office for a grant proposal or to the editor for a journal submission. Upholding fairness as a research value, as discussed in Chapter 2, requires that reviewers be aware of their own biases so as to avoid critiques that are motivated by a desire to defend their own work. The program officer or editor has the responsibility to decide whether a bias or conflict of interest affects a potential reviewer's eligibility.

Reviewers also need to uphold the confidentiality of the review process by

not sharing materials or ideas from grants or manuscripts under review. Appropriating ideas from grants or manuscripts under review is a form of plagiarism.

**Best Practice R-6. Research Compliance. Understand and comply with relevant institutional and governmental regulations governing research, including those specific to a given discipline or field.**

Research often involves risks to human subjects and animals, to those in the lab, or to those in the buildings where the research takes place. Because research has a potential for harm, it is regulated by local, state, or federal laws, and human and animal studies are governed by Institutional Review Board and Institutional Animal Care and Use Committee rules, respectively, and regulations imposed by the federal government. Failure to comply with governing rules and regulations can lead to civil—or in some cases criminal—penalties for researchers. Moreover, compliance failures undermine public confidence in the researcher, the institution, the field, and the broader research enterprise.

Researchers have the responsibility to determine what the governing rules are for a designed experiment before the work is conducted. Most institutions have offices that specialize in safety, human experiments, and animal use. These offices should be consulted fully to ensure safety—of the researchers and participants in the experiment or the larger community—and that all governing rules and regulations are satisfied. In some fields, researchers also need to be aware of the risks inherent in doing science, understand the possibilities of harmful consequences that could arise accidentally or through misuse, and take steps to reduce those risks as much as possible.

Finally, researchers need to disclose personal financial interests that might reasonably appear to be related to the research for review by institutional officials at the appropriate time. In many cases, the conflict can be managed through the actions of the researchers involved and through oversight. In some cases, the conflict may not be manageable and must be eliminated or the project may have to be abandoned. Personal financial interests related to the research may have the effect of undermining a reader's view of the credibility of the results, but honesty and objectivity require that they be listed so that others can draw conclusions about the possible effects.

A best practices checklist for researchers is provided in Box 9-1.

## Research Institutions

As the employers of researchers and the institutional stewards of financial and other resources that support research, universities and other research institutions in the United States have a number of responsibilities (both formal and informal) for ensuring integrity. According to the Institute of Medicine and the National Research Council, "Each research institution should develop and implement a comprehensive program designed to promote integrity in research, using

**BOX 9-1**
**Best Practices Checklist for Researchers**

**Research Integrity**
- Maintain high standards in own work.
- Understand policies.
- Raise questions and problems promptly and professionally.
- Strive to be a generous and collegial colleague.

**Data Handling**
- Develop data management and sharing plan at the outset of a project.
- Incorporate appropriate data management expertise in the project team.
- Understand and follow data collection, management, and sharing standards, policies, and regulations of the discipline, institution, funder, journal, and relevant government agencies.

**Authorship and Communication**
- Ensure that general and disciplinary standards are followed for research publications.
- Acknowledge the roles and contributions of authors.
- Be transparent when communicating with all audiences.

**Mentoring and Supervision**
- Model and instruct on research best practices.
- Regularly check work of subordinates and ensure adherence to best practices.
- Clarify expectations.

**Peer Review**
- Provide complete and timely review.
- Maintain confidentiality.
- Disclose conflicts, and eliminate or manage them as appropriate.

**Research Compliance**
- Protect human subjects and laboratory animals.
- Follow environmental and other safety regulations.
- Do not engage in misuse.
- Disclose and manage conflicts of interest.

multiple approaches adapted to the specific environments within each institution." (IOM-NRC, 2002) Specific responsibilities include the maintenance of policies and procedures to investigate and address research misconduct—including the responsibility to notify the appropriate federal agency of misconduct investigations involving that agency's funds—and the provision of educational and training programs for students and faculty to raise awareness of research integrity (IOM-NRC, 2002; NAS-NAE-IOM, 1992; NSF-OIG, 2013; OSTP, 2000).

In addition, research institutions carry a range of research-related legal and regulatory compliance responsibilities, such as administering regulations governing research on human subjects and laboratory animals; acting as stewards, as

required, of data from federally funded research (see NAS-NAE-IOM, 2009a); enforcing environmental and hazardous substance regulations; ensuring proper financial accounting of research funds; and implementing general workplace laws and regulations in areas such as discrimination and harassment. The challenges presented by these myriad, often overlapping regulations are many. Institutional leadership must take a role in seeking a responsible compliance environment that is designed to facilitate and support a quality working and learning environment for all.

Some specific policies and practices of research institutions may differ according to whether they are controlled and operated by public or private universities, other nonprofit entities, for-profit companies, or government bodies. Presentations to the committee by corporate representatives indicated that some multinational companies take a very thorough and systematic approach to training and mentoring young researchers (Williams, 2012).

As experience has accumulated over the past several decades, new perspectives have appeared regarding how research institutions can best foster research integrity. For example, the practice of assessing the climate for research integrity in an institution has emerged and is becoming more widely adopted, and its benefits are becoming more clearly understood (CGS, 2012; IOM-NRC, 2002). Around the world, more attention is being paid to the role of universities and research institutions in ensuring integrity (ESF-ALLEA, 2011; UUK, 2012). The responsibilities of universities and research institutions may change over time due to the challenges raised by new technologies and collaborations (IOM, 2009, 2012).

**Best Practice I-1: Management. Integrate research integrity considerations into overall approaches to research, education, and institutional management.**

Changes in the funding, structure, and organization of research in the United States and the possible effects of these changes on the incentives of researchers to uphold best practices are discussed in several places in this report. In fulfilling their responsibilities to create an environment where the fundamental values of research are valued and reinforced, institutions need to consider organizational and management issues that have not traditionally been associated with research integrity and have not been traditionally seen as organizational responsibilities. In this regard, institutional leaders and others with research administration responsibilities need to demonstrate through their approach to oversight and implementation of policies that fostering research integrity is a central priority that supports the quality of research. It would be a mistake for institutional and faculty leaders to observe that the institution has basic policies and administrative procedures in place and assume that research integrity issues do not require their attention.

While this is a broad exhortation compared with other best practices presented here, the committee identified several areas for particular focus during the course of the study. To begin, institutions should explicitly evaluate mentoring as

part of their evaluation of faculty. Mentoring and supervision of young researchers at U.S. institutions needs systematic attention and improvement. A review of closed Office of Research Integrity (ORI) cases found that almost three-quarters of supervisors had not reviewed source data with trainees who committed misconduct and two-thirds had not set standards for responsible conduct (Wright et al., 2008). Another recent survey of research faculty found that less than a quarter have had opportunities to participate in faculty training to be a better mentor, advisor, or research teacher, and about one-third of faculty did not or could not remember whether they had guidelines related to their responsibilities to PhD students (Titus and Ballou, 2014). Recent work by the InterAcademy Partnership indicates that the need for improved mentoring of young researchers is a global issue (IAP, 2016).

Another imperative is to regularly communicate relevant institutional policies—such as the definition of research misconduct—as well as the rights and responsibilities of researchers directly to young researchers. Compacts between institutions and postdocs, students, and faculty are one mechanism for such communication. The American Association of Medical Colleges has developed several sample compacts, including one between graduate students and their research advisors and one between postdocs and their mentors (AAMC, 2006, 2008). These are documents of several pages that include bullet points outlining the responsibilities of both parties, such as the responsibility of graduate students to seek regular feedback and the responsibility of graduate advisors not to require students to perform duties unrelated to training and professional development. A particularly important and sometimes vulnerable group is postdocs (Phillips, 2012). Postdocs are formally trainees but are often called upon to be mentors of students or younger postdocs. A 2005 survey of postdocs found that less than half of respondents were aware of institutional policies toward determining authorship, defining misconduct, resolving grievances, or determining the ownership of intellectual property (Davis, 2005).

A related responsibility is for institutions to collect data on career outcomes for recent science and engineering graduate cohorts and postdocs and to provide these data to incoming students and trainees at the front end of their training programs so they are better informed. Providing this information is one indication that the institutions have the students' best interests at heart. To the extent that students have a realistic perspective of their career prospects and the likelihood of being able to pursue research as a career, they will be better equipped to make decisions about how to proceed with their graduate training

Further, institutions might benefit from keeping track of such organizational and funding issues as the number and proportion of soft-money positions in various departments, as well as trends. As explored elsewhere in the report, the combination of increasing emphasis on soft-money positions and declining success rates for grant applications at agencies such as the National Institutes of Health may have a negative impact on researcher incentives to uphold high standards.

Finally, the committee has noted a trend toward institutions and researchers undertaking more aggressive public relations efforts on behalf of their research activities. Institutions and researchers should impose careful quality control on such efforts. One recent study indicates that the quality of media reporting on discoveries is directly related to the quality of press releases (Schwartz et al., 2011). Well-known cases over the years of aggressively promoted results that turned out to be based on fabricated data, such as the Hwang stem cell case, or were otherwise irreproducible, such as the Fleischmanm-Pons "cold fusion" discovery, provide cautionary tales (Appendix D; Goodstein, 2010). Overhyping may ultimately be both a cause and a consequence of a "winner take all" culture in research where disincentives to cutting corners, or even worse behaviors, are weakened over time (Freeman and Gelber, 2006; Freeman et al., 2001a,b). It may also damage public trust in researchers and in the research enterprise.

**Best Practice I-2: Assessment. Perform regular assessments of the climate for research integrity at the institutional and department levels and address weaknesses that are identified.**

A baseline expectation is that institutions should create a climate for research integrity and institute supportive policies and practices. The 2002 report *Integrity in Scientific Research* explains that research organizations "engage in activities that help establish an internal climate and organizational culture that are either supportive of or ambivalent toward the responsible conduct of research" (IOM-NRC, 2002). That report recommended that institutions utilize ongoing self-assessment and peer review in order to evaluate their climate for research integrity and guide continuous improvement. At that time, instruments for that purpose had not been developed.

In recent years, an instrument to assess the organizational climate for research integrity has been developed and validated (Crain et al., 2013; Martinson et al., 2013). A recent Council of Graduate Schools (CGS) project worked with a group of universities to integrate "research ethics and the responsible conduct of research (RCR) into graduate education" (CGS, 2012). The participating universities utilized climate assessment as an important tool to identify areas for improvement and to track progress. One participating institution reports that the data produced by the assessment tool helped efforts to improve research integrity approaches gain traction among the faculty (May, 2013).

Institutions can also assess the effectiveness of their own efforts to promote research integrity. Are allegations or concerns addressed in an appropriate and timely way? Are policies related to transparency and data sharing well understood and followed?

Strengthening education and training in the responsible conduct of research, discussed below, is an important approach to addressing issues uncovered in assessment exercises and improving local research climates. As illustrated by several of the cases discussed in Appendix D and in other parts of the report, if

detrimental research practices are tolerated at the laboratory or department level, it can lead to a vicious circle where young researchers perpetuate these practices in the belief that they are behaving appropriately. In response, institutions might look for other proactive approaches such as placing succinct posters on bulletin boards to encourage best practices. ORI has produced an infographic on how research supervisors can foster integrity that provides an example of the sorts of information that might be communicated (ORI, 2016). The Singapore Statement on Research Integrity (2010) produced by the Second World Conference on Research Integrity is also available as a single-page pdf. Such posters would perhaps be more effective if they were locally produced by labs or departments.

**Best Practice I-3: Performing Research Misconduct Investigations. Perform regular inventories of institutional policies, procedures, and capabilities for investigating and addressing research misconduct and address weaknesses that are identified.**

Universities and other research institutions are responsible for undertaking fair, thorough, and timely investigations into allegations of research misconduct. A comprehensive assessment of how U.S. research institutions are performing in the area of addressing research misconduct is not possible, because most investigation results and reports are never made public due to confidentiality rules. Over the course of the study, experts who briefed the committee pointed to considerable unevenness in the capabilities of universities to investigate and address allegations of research misconduct (Garfinkel, 2012). In addition, the examples described in other parts of the report, particularly Chapter 7 and Appendix D, illustrate that even the most highly regarded institutions can fail in the performance of basic tasks, such as following appropriate investigation procedures, ensuring that internal committees have the right knowledge and expertise, and ensuring that investigation processes avoid the pitfalls that can result from institutional conflicts of interest.

Regular inventories of institutional policies, procedures, and capabilities can help to ensure that the minimum requirements needed to comply with existing regulations are met, but universities should aim for more than compliance. The requirements of ORI and the National Science Foundation (NSF) should be a floor, not a ceiling.

Ensuring that institutions have the appropriate policies and resources in place to address research misconduct allegations starts with the support and involvement of institutional leaders. Often, concerns can be addressed and questions can be answered at an early stage, obviating the need for formal investigations (Gunsalus, 1998b).

Elements that should be part of institutional capabilities include a trained Research Integrity Officer or other professional who can act on allegations, involvement of the institution's general counsel's office, clear policies and procedures that are understood and followed, and support from institutional leadership. In

research universities, faculty leaders play a critical role in the effective communication and implementation of these policies and procedures. Institutions should also protect good-faith whistleblowers and prevent negative career consequences for young researchers who become whistleblowers. This demonstrates the institution's moral commitment to its students and employees. As illustrated by the Goodwin case, young researchers who do the right thing by raising concerns or making allegations against superiors may find that their research careers are effectively over, even when they uncover misconduct.

Maintaining confidentiality during an investigation, protecting the accused, and minimizing the negative consequences of investigations for those who are cleared are also essential. Institutions need to communicate with federal agencies such as ORI and the NSF Office of Inspector General, sponsors, and journals, as appropriate, to ensure that these entities can fulfill their responsibilities related to the stewardship of funds and correcting the research record.

Institutions also need to have policies and mechanisms in place that allow them to call in external sources of expertise, particularly when their financial, reputational, or other interests may be affected by an allegation. Incorporating external members on the institutional committees that undertake research misconduct investigations is one mechanism for accomplishing this. In some particularly serious or problematic cases, an institution may decide that all members of such a committee should come from outside the institution, although considerations of logistics and cost would make it difficult to institute this as a normal practice. The University of Illinois requires that all investigation committees should include at least one external member (University of Illinois, 2009). In addition, institutions may ask external experts to review the mission statements of investigation committees at the start of the process and the draft reports of committees to help ensure that the appropriate questions and issues are addressed. It is not clear how common external review is currently.

Regular evaluations of capabilities, incorporating perspectives external to the institution, can also help institutions improve their systems and processes over time. For example, in addition to designated institutional points of contact for allegations of misconduct, such as Research Integrity Officers, some institutions have found additional resources, such as ombudsmen and hotlines, to be helpful. In managing a system with multiple entry points, it is necessary to clearly define roles and coordinate responses so that those who are bringing their concerns to the institution do not receive incorrect or conflicting advice. Mediation mechanisms can be put in place for disputes that arise between colleagues or between subordinates and superiors. Ideally, enhanced communication and related interventions will allow many issues and concerns to be addressed before research misconduct occurs. Ensuring that this information is widely disseminated through posting on bulletin boards in labs and through other mechanisms is also important.

**Best Practice I-4: Training and Education. Strive for continuous improvement in RCR training and education.**

The development of RCR training and education programs and related issues—including funder mandates, content, delivery mechanisms, and assessment—are covered in detail in Chapter 10. The 1992 report *Responsible Science* noted that institutional RCR education programs were not very common at that time and that the research enterprise was ambivalent about such programs (NAS-NAE-IOM, 1992). Although there is still much to be learned about the effectiveness of particular educational approaches, recognition that institutions have clear responsibilities has grown over time, both in the United States and around the world. The report *Integrity in Scientific Research* recommended that "institutions should implement effective educational programs that enhance the responsible conduct of research" (IOM-NRC, 2002). The *Australian Code for the Responsible Conduct of Research* states that

> Each institution must provide induction and training for all research trainees. This training should cover research ethics, occupational health and safety, and environmental protection, as well as technical matters appropriate to the discipline. (NHMRC-ARC-UA, 2007)

As is the case with institutional policies and resources to address allegations of research misconduct, the formal requirements of funders should constitute the floor, not the ceiling, for institutional efforts. NIH mandates participation in RCR education for all persons receiving NIH support. This requirement includes instruction in nine core areas: (1) data acquisition, management, sharing, and ownership; (2) mentor/trainee responsibilities; (3) publication practices and responsible authorship; (4) peer review; (5) collaborative science; (6) human subjects; (7) research involving animals; (8) research misconduct; and (9) conflict of interest and commitment (Steneck, 2004). A 2009 update on the Requirement for Instruction in the Responsible Conduct of Research requires RCR training to be provided in person, noting that online instruction is a helpful supplement but is insufficient as the sole provider of RCR training (NIH, 2009). The guidance suggests at least a semester-long series of RCR instruction from faculty on a rotating basis to ensure full faculty participation and that instruction recur through the different levels of a scientist's career (NIH, 2009). The CGS project discussed below produced a number of possible approaches for institutions aiming to improve RCR education, such as engaging faculty in developing discipline-specific content, holding lunchtime workshops for graduate students, integrating RCR content into courses, and developing courses that escalate in complexity (CGS, 2012). The *Integrity in Scientific Research* report also recommends RCR instruction be provided by "faculty who are actively engaged in research related to that of the trainees" (IOM-NRC, 2002). The CGS project made recommendations for institutional leaders to demonstrate engagement in RCR education through public

endorsement from the university president and by assembling a steering committee of institutional leaders and a project director to oversee a plan to integrate RCR education into the curriculum (CGS, 2012).

Institutions can participate in and take advantage of other RCR education development efforts. Recently, RCR training has shifted emphasis from the traditional focus on imparting knowledge, specifically of regulations and compliance requirements, toward the potential value of imparting skills in ethical decision making (see Appendix C). The effectiveness of techniques such as team-based learning is also being explored (McCormack and Garvan, 2014). An organization involved in RCR is the National Postdoctoral Association, which oversaw a project aimed at developing RCR educational approaches specifically for postdocs (NPA, 2013).

Box 9-2 provides a best practices checklist for research institutions.

### Journals and Other Scholarly Communicators

This section and the associated practices are addressed to journals—editors, governing bodies, and publishers—and other individuals and groups involved with scientific publishing and other forms of scholarly communication, including university librarians, digital archivists, and academic presses.

The basics of responsible publishing include ensuring that a journal's existing rules and guidelines have been followed, such as those related to data sharing and research involving human subjects (Gustafsson et al., 2006). Editors are also responsible for the scientific quality of the journal. Journals should clearly articulate their publication criteria and evaluate submissions based on those criteria. They should provide the authors of proposed publications with a fair and full account of reviewers' comments and ensure transparent communication in the event of disputes, questions, or difficulties in the publication process. Journals should make their principles and processes visible to authors, readers, librarians, and peer reviewers. As an example, publishers should disclose sources of funding or other issues that may affect the choice of work to disseminate.

The 1992 report *Responsible Science* mentions scientific journals and editors and contains a general recommendation that journals and societies support research integrity. Journal concerns and responsibilities related to research integrity have grown and shifted in recent years, as article retractions have increased, a series of high-profile cases of fabricated research published in several high-profile journals has come to light, and relatively new challenges such as image manipulation have prompted journals to develop new policies and approaches. The fact that detecting fabrication often requires specialized technical and analytical tools makes it unlikely that it will be uncovered in the normal peer review process (i.e., before publication).

Although it is sometimes assumed that journal peer review processes are or should be effective mechanisms for uncovering fabricated data and other re-

**BOX 9-2**
**Best Practices Checklist for Research Institutions**

**Research Integrity and Institutional Management**

- Demonstrate that fostering research integrity is a central priority at all levels, including for faculty and institutional leaders.
- Provide training to faculty in effective mentoring and include mentoring as a criterion for hiring and promotion.
- Communicate rights and responsibilities to students, faculty, postdocs, and others engaged in research (e.g., through the use of compacts or other mechanisms).
- Collect and disseminate data on the career prospects of graduate students and postdocs.
- Consider implications for research integrity when making larger management decisions—(e.g., the number and proportion of soft-money positions).
- Do not exaggerate research results in institutional communications.

**Climate Assessment**

- Gather data on institutional climate related to research integrity.
- Share data across graduate departments.
- Share practices of strong departments and address shortcomings of weak departments.

**Performing Research Misconduct Investigations**

- Meet formal compliance responsibilities by ensuring that policies and capabilities for performing fair, thorough, and timely investigations of research misconduct allegations are in place.
- Have multiple entry points to raise questions about possible misconduct.
- Use checks and balances to guard against institutional conflicts.
- Involve legal counsel.
- Incorporate external perspectives when appropriate.
- Protect whistleblowers during investigations and mitigate negative consequences on their careers afterwards.
- Take "after action steps" to ensure that papers are retracted.

**RCR Training and Education**

- Engage faculty.
- Make federal requirements a floor, not a ceiling.

search misconduct, history and recent experience indicate that this is not the case (Ioannidis, 2012; Stroebe et al., 2012). Most misconduct is uncovered through revelations by whistleblowers or by other scientists who have tried and failed to replicate fabricated research.

Over the years, a number of individual journals and publishing groups, journal associations, and other groups have developed ethical codes and good practice guidelines for scientific publishing (COPE, 2011; CSE, 2012b; ICMJE, 2013; SfN, 2010). Some publication executives and boards regard the Committee

on Publication Ethics (COPE) principles and recommendations as directive and more or less adhere to them. Others regard them as informative and suggestive while holding independent views on responsible publishing that occasionally vary from COPE's advice. COPE promulgates a mandatory code of conduct for journal editors and a more aspirational set of best practices. COPE has also published a number of guidelines and monographs intended to assist editors and publishers in the course of their work.

Digital innovation has been a major source of disruption in science, engineering, technology, and medical research and publishing, and this has implications for responsible research. Predicting the directions and extent of progress in information technologies is difficult, yet principles and best practices in publishing should be flexible enough to be applied as innovations in research practice arise. The Society for Neuroscience's recently revised ethics policy and guidelines for responsible conduct in scientific publishing are useful examples (SfN, 2010). The set of guidelines put forward for authors is notable for the detailed specifications given for describing the intellectual contribution of authors.

Some journals have introduced technical checks to detect plagiarism and image manipulation. These tools have been useful in detecting misconduct and detrimental practices in proposed papers. In addition, a recent trend among biomedical journals has been to hire ethics officers. It should be noted that these sorts of steps contribute to rising costs that are passed on to university libraries, other subscribers, and, in the "open access" arena, the authors of research. Still, these costs need to be balanced against the costs incurred in editorial time when a journal has to retract a paper.

**Best Practice J-1: Practicing Transparency. Practice transparency in journal policies and practices related to research integrity, including publication of retractions and corrections and the reasons for them.**

Openness is fundamental to the success of the entire chain of processes and relationships involved in scholarly communication. This principle translates directly into best practices in publishing, with just a few exceptions. The one obvious exception is that of peer review, in which the identity of peer reviewers has traditionally been hidden so that undue influence on reviewers is minimized, pre- or postpublication, thus creating an environment enabling direct and frank critical commentary for authors and editors by reviewers. As discussed in Chapter 3, improving peer review policies and practices and considering other models—such as unblinded review—are issues currently facing journals and disciplines.

Following this best practice begins with maintaining an up-to-date set of author instructions, as well as ethical policies for authors, reviewers, and editors. The policies should include procedures to be followed when allegations of misconduct arise. Journals should communicate retractions (including the reasons for retractions or why a reason cannot be provided), corrections, clarifications, and apologies promptly and openly to ensure that the published record of research is

as free of bias, error, and falsehoods as possible. New means of electronic communication provide new and potentially powerful ways of correcting the research literature. There is great value in putting retractions in the place of the target article *and* in tables of contents. Metadata—which is information about a dataset embedded within it—associating each with the target article should be included for ongoing observation and analysis.

In addition, data and code that support an article should be published with the article (or chapter or book) or made otherwise available (e.g., through linking) in its original position in an issue (or edition) as well as a separate issue- or title-level section with its own explicit entry in the table of contents. Publishers and editors should provide for postpublication review and commentary attached to scientific, technical, and medical articles. Such commentary can be helpful in uncovering problems with published work and in exploring promising areas for research that would confirm or extend the reported results.

Journals should have policies in place to prevent conflicts of interest on the part of editorial staff from affecting editorial decisions. One way of handling this would be for editorial staff to provide conflicts of interest in narrative form in articles and as metadata for systematic observation and analysis. Alternatively, the journal might define what constitutes a conflict of interest for any editor, and then state that if an editor has a conflict of interest with any of the authors of a paper, he or she is excluded from handling the paper. Journals would have on hand declarations from their editors that are updated annually or more often as circumstances change. Addressing conflicts of interest of other participants in the publication process is covered below.

Throughout the publishing process, journals should negotiate fairly and as transparently as possible in author, author-reviewer, and author-reader disputes.

While not as directly supportive of research integrity as the other steps outlined above, journals contribute to the effective functioning of the research enterprise by providing open access to publications, perhaps after an embargo period so as not to interfere with a publisher's business viability.

**Best Practice J-2: Requiring Openness. Require openness from authors regarding public access to data, code, and other information necessary to verify or reproduce reported results. Require openness from authors and peer reviewers regarding funding sources and conflicts of interest.**

As described in other parts of this report, including Chapter 7, requiring authors to share data and code for purposes of verification, replication, and reuse is an important step that the research enterprise can take to help ensure research integrity. Journals are in a powerful position to implement this step, and some are developing new policies and procedures aimed at ensuring access to data and code (*Nature*, 2013). Although making data available with the article is the traditional approach in many disciplines, linking to a specialized database or repository will likely be the preferred way to provide access to data in most cases.

One example of efforts to expand the availability of data is a 2016 proposal by the International Committee of Medical Journal Editors that in order for an article to be considered for publication authors should be required to commit to publish "deidentified individual-patient data underlying the results" of clinical trial research within 6 months of the corresponding article for reproducibility purposes (Taichman et al., 2016).

The data to be made available should include outlier data and negative results if appropriate. Alterations to images should be specified. In cases where regulatory, legal, or technological constraints prevent authors from providing full access to data, an explanation should be published along with the paper.

Journals should work with sponsors, authors, and research institutions to ensure long-term access to data, code, and other information supplementary to the article. Archiving of articles and supplementary information by third parties is the ultimate goal, although securing the necessary resources and developing the appropriate mechanisms remain challenging tasks in some fields and disciplines.

It is also important for full method descriptions to be included in every publication. Currently, references to method sections in previously published work are common in some fields, but this may cause ambiguity as to what was actually done. With the availability of electronic supplements, there is no reason why full methods cannot be included, even if this means reprinting what the same author published previously. Good practice should not be discouraged by concerns about self-duplication if this increases transparency and reduces ambiguity.

Financial conflicts of interests, other relevant financial relationships, and relevant nonfinancial interests should be identified by all authors and included in print and as metadata (PLOS Medicine Editors, 2008). For example, "publishing relevant competing interests for all contributors and publishing corrections if competing interests are revealed after publication" is a best practice listed in COPE's guidelines (COPE, 2011). This disclosure should include an explicit citation of support from funders, whether corporate or not for profit.

Journals should also take steps to safeguard the integrity of the peer review process. COPE's guidelines for peer reviewers include submitting a declaration of potential competing interests, respecting the confidentiality of the process, and not intentionally delaying the process (Hames, 2013). Journals might ask reviewers to explicitly commit to these guidelines by signing a statement.

**Best Practice J-3: Authorship Contributions. Require that the contributions and roles of all authors be described.[2]**

[2] In Recommendation Five, this report calls for the development and adoption of authorship standards and suggests a framework that if adopted would formally codify the requirement that the roles of authors be disclosed across all fields and disciplines. See Chapter 8 for the rationale underlying the recommendation and Chapter 11 for the recommendation text.

Article authors are the researchers who have contributed significantly to the article and are listed in the article byline. Authorship determines who receives credit for the work and fixes responsibility if or when mistakes or misconduct is uncovered. While guidance on authorship is provided by journals, institutions, societies, and other groups, specific practices vary by discipline. Although detrimental authorship practices other than plagiarism have not been included in the U.S. government's definition of research misconduct, practices such as honorary authorship and unacknowledged ghost authorship, as well as authorship disputes, pose challenges to research integrity. The Council of Science Editors points out that "problems with authorship are not uncommon and can threaten the integrity of scientific research" (CSE, 2012b). A recent review of research on authorship across all fields found that 29 percent of researchers in several separate studies reported that they or others they know had experiences involving the misuse of authorship (this figure could be inflated by multiple reports of the same behavior in some of the reviewed studies) (Marušić et al., 2011).

In an environment of increasing collaboration across institutions and borders, it may be more difficult to determine who is responsible for mistakes or fabricated work. In some cases of fabricated or falsified research, senior researchers have claimed that they were merely honorary authors and therefore were not responsible for the integrity of the reported work.

These issues pose challenges to journals, which have responded by paying increasing attention to authorship. One journal practice that has become fairly widespread is to require authors to describe their individual contributions, which are published in a designated place in the article. Journals such as the *Lancet* began adopting this practice in the 1990s (Yank and Rennie, 1999). The Nature Publishing Group journals, which had requested that authors provide contribution disclosures beginning in 1999, made them mandatory in 2009 (*Nature*, 2009). At the same time, *Nature* had considered requiring corresponding authors to sign a statement that they had taken some integrity assurance steps, but there was significant skepticism about this proposal.

Most current contribution disclosures tend to be fairly broad. For example, the *Proceedings of the National Academy of Sciences* provides an example list of contributions that includes research design, research performance, contribution of new reagents or analytic tools, data analysis, and writing (PNAS, 2013). Advances in technology hold out the possibility that such contribution disclosures can become more detailed and useful in the future, providing the underlying tools for researchers to maintain up-to-date, verified accounts of their work (Frische, 2012).

For now, journals should require contribution disclosures at as detailed a level as practical and be open to adjusting these requirements as technologies and tools evolve. For peer-reviewed papers, all authors should be identified along with the sources of funding for their work. To avoid questions of duplication, previously published materials should be identified and cited.

**Best Practice J-4: Training and Education. Facilitate regular training and education in responsible publishing policies and best practices for editors, reviewers, and authors.**

Best practices for research institutions and mentors in RCR training and education are described above. Journals can play an important role in focused areas of RCR education as well. It is particularly important for editors to be knowledgeable about responsible publishing practices, requirements that need to be communicated to authors and reviewers, and what to do if problems arise. Some aspects of responsible writing, reviewing, and editing may not be covered in RCR training provided to graduate students. A recent review indicates that many writers, reviewers, and editors lack the necessary training to play their roles effectively, but little is known about the availability and effectiveness of such training (Galipeau et al., 2013). The Council of Science Editors, which has provided training for editors for some time, recently launched a certificate program in scholarly publication management (CSE, 2012a). A 2006 paper recommended that an international online training and accreditation program for peer reviewers should be established (Benos et al., 2007).

Journals have varied capabilities and resources to encourage training or to undertake their own educational programs. They should take what steps are appropriate to their own circumstances to help ensure that authors, reviewers, and editors are well prepared to perform their tasks.

**Best Practice J-5: Collaboration. Work with other journals to develop common approaches and tools to foster research integrity.**

As described elsewhere in this section, the work of groups such as the Committee on Publication Ethics, International Committee of Medical Journal Editors, and Council of Science Editors has been of great value to the research enterprise in developing policies, tools, and approaches to ensure research integrity. While individual journals and other scholarly communicators need to maintain the independence to adopt policies and practices that are appropriate to their circumstances, continued collective efforts by journals can contribute to improvements in standards and practices across the enterprise. Uniform policies reinforce the norms of research integrity.

Box 9-3 provides a best practices checklist for journals and other scholarly communicators.

### Research Sponsors and Users of Research Results

Sponsors and users of research occupy particularly important positions in the research enterprise. In general, researchers and research institutions rely on funding from government and private-sector sponsors such as industry and foundations to perform their work. The incentive structures created by sponsors

**BOX 9-3**
**Best Practices Checklist for Journals**

**Practicing Transparency**
- Adopt up-to-date policies and instructions.
- Publish retractions/corrections and reasons in articles, in tables of contents, and as metadata in a timely fashion.
- Provide a link to data and code that support articles, and facilitate long-term access.
- Require full descriptions of methods in method sections or electronic supplements.
- Provide for postpublication review and commentary.
- Be transparent in negotiating with authors and in adjudicating disputes.
- Establish a conflict-of-interest policy covering editorial staff.
- Provide open access consistent with business viability.

**Adopt Policies That Ensure Openness Regarding:**
- Data, code, and records of any image alterations.
- Author funding and conflicts of interest.
- Peer reviewer conflicts of interest.

**Author Contributions**
- Describe author roles.

**Training and Education**
- Facilitate training for editors, reviewers, and authors.

**Collaboration**
- Participate in science, engineering, technology, and medical publishing efforts to develop tools and approaches to foster integrity.

can have a significant influence on the motivations and behaviors of researchers and institutions. The changing environment for research funding and the resulting pressures on researchers are described in Chapter 3 and Chapter 6. While specific recommendations to sponsors are developed in Chapter 11, this section identifies several specific best practices that research sponsors and users of research results can adopt to ensure research integrity.

The 1992 report *Responsible Science* recommended several roles for government research sponsors related to integrity, including adopting a common framework of definitions of research misconduct and common policies, adopting policies and procedures that ensure appropriate and prompt responses to allegations of misconduct, and providing support for institutional efforts to discourage questionable research practices (NAS-NAE-IOM, 1992). The 2002 report *Integrity in Scientific Research* recommended that research sponsors support work to increase understanding of the factors that influence research integrity, including monitoring and assessing those factors (IOM-NRC, 2002). As discussed in

Chapter 6, the Office of Research Integrity and the National Science Foundation maintain programs to support such research.

U.S. government research sponsors such as the National Institutes of Health and the National Science Foundation have imposed several mandates and other regulatory requirements on research institutions and researchers over the past several decades covering RCR education and training. The Office of Research Integrity also requires institutions to file an assurance that they have developed and will comply with policies for addressing allegations of misconduct in Public Health Service–sponsored research that meet Public Health Service policies.

The need for research sponsors to take an active role in fostering research integrity is becoming more recognized around the world. The Irish Council for Bioethics report *Recommendations for Promoting Research Integrity* (ICB, 2010) provides a useful overview of various approaches. The Global Research Council's *Statement of Principles on Research Integrity* is a succinct list of funding agency responsibilities that includes promotion of education, leading by example, and conditioning support on upholding research integrity (GRC, 2013). The Inter-Academy Council and InterAcademy Panel (IAC-IAP, 2012) have also described the responsibilities of funding agencies in *Responsible Conduct in the Global Research Enterprise: A Policy Report*.

**Best Practice RS-1. Research Integrity and Quality. Align funding and regulatory policies with the promotion of research integrity and research quality.**

Aligning funding and regulatory policies with the promotion of research integrity and research quality has several distinct aspects. For example, as described in Chapter 4, some funding agencies and regulatory bodies maintain policies on research misconduct and exercise oversight over how institutions address allegations of misconduct. Private foundations such as the Howard Hughes Medical Institute also have research misconduct policies (HHMI, 2007). As discussed in Chapter 9, agencies require grantee institutions to provide RCR education. Funders that play these roles should ensure that their policies are clear and implemented consistently. Additional commentary on the policies and practices of U.S. government agencies is provided in Chapter 7 in support of the committee's recommendations in this area.

A second aspect of aligning policies and practices with the promotion of research integrity is to increase awareness of how funding policies affect research integrity and to make adjustments when possible and necessary. This may involve support for research that illuminates issues related to research integrity. For example, in recent years the Office of Research Integrity has responded to evidence that the institutional environment has a major impact on research integrity by supporting efforts to study, assess, and strengthen those environments. Some policy initiatives might be based on direct understanding of a situation rather than the results of sponsored research—ORI has also sought to address unevenness

in institutional capacity to respond to allegations of misconduct by supporting professional training for research integrity officers.

A recent international report has pointed out that funders have a responsibility to ensure that funding policies not cause researchers and research institutions to emphasize quantity over quality (IAC-IAP, 2012). Chapter 6 explores whether changes in the level and structure of research funding might be associated with detrimental research practices or misconduct. As explained there, this is a complex issue. Evaluating the extent of possible problems and recommending solutions are beyond the scope of this committee's task. Nevertheless, agencies may already be collecting relevant data on how changes in funding and organization are affecting research environments (NIH, 2012a). Sponsors should look for opportunities to develop evidence on possible impacts of funding policies on the researchers and institutions that are supported, including impacts on integrity, and take appropriate actions. One example is the NIH policy that limits the number of publications that can be listed in the biosketch submitted in grant and cooperative agreement applications, which may help reduce incentives for researchers to maximize the number of publications (NIH, 2014).

Finally, research funders can take steps to coordinate and harmonize their activities within their own domestic contexts as well as internationally. Examples of international cooperation include NSF's participation in the Global Research Council and Organisation for Economic Co-operation and Development Working Group activities to develop common approaches to dealing with research integrity issues across member countries (GRC, 2013; OECD, 2009, 2007). The Fogarty International Center, part of NIH, supports capacity building in bioethics and research integrity in the developing world.

**Best Practice RS-2. Data and Code. Promote access to data and code underlying publicly reported results.**

The importance of ensuring access to data and code for research integrity and quality is covered above with reference to journal practices and policies. Funders have important roles to play as well. The America COMPETES Reauthorization of 2010 called on federal agencies to ensure access to publications and data resulting from work that they support, and the Office of Science and Technology Policy began working with agencies on implementing the legislation in early 2013 (Holdren, 2013). Federal sponsors can also play a role in providing resources to cover the costs borne by researchers and institutions in making data and code available. Funders will play a critical role in supporting the development of necessary infrastructure, such as data and sample repositories, efforts to develop metadata standards, and the development of applications that facilitate the direct deposit of data to the repositories complete with the metadata. Without those efforts and tools, compliance for data deposition will be low, and the ability of others to use the data for reproducibility will be hampered.

Industry research sponsors also have important contributions to make in

this area. Clinical trial data constitute a prominent specific example. Over the years, the share of clinical trials funded by industry has grown (Buchkowsky and Jewesson, 2004). At the same time, pressure has grown to make the clinical trial process more transparent through mechanisms such as public registration of all trials and encouraging the release of all results, including negative results. A recent report states that there are "compelling justifications for sharing clinical trial data to benefit society and future patients" (IOM, 2015). There is a need to ensure that data sharing is done responsibly and protects privacy. Lack of timely reporting of clinical trials is not solely or even primarily an issue in industry-performed or industry-sponsored work; clinical trials performed at academic medical centers and sponsored by federal agencies and other nonindustry sources also need to improve their practices (Chen et al., 2016). Still, since clinical trials are an important component of industry-sponsored research that is published in peer-reviewed journals, industry sponsors can make an important contribution by registering all of their trials, reporting all results in a timely way, and sharing data responsibly.

In September 2016, NIH issued a final policy to promote broad and responsible dissemination of information from NIH-funded clinical trials through ClinicalTrials.gov. Under this policy, every clinical trial funded in whole or in part by NIH is expected to be registered on ClinicalTrials.gov and have summary results information submitted and posted in a timely manner, whether subject to section 402(j) of the Public Health Service Act or not (NIH, 2016).

**Best Practice RS-3: Utilizing Research. Practice impartiality and transparency in utilizing research for the development of policy and regulations.**

As discussed in Chapter 3, scientific evidence and inputs are increasingly important to numerous areas of policy making—public health, environmental protection, economic development, criminology, food safety, education, and many other areas. The interpretation of research results is a central part of many contentious policy debates, which often feature accusations that science is being manipulated or distorted by powerful interests.

One recent report identifies the five "tasks" that science has in relation to policy: "(1) identify problems, such as endangered species, obesity, unemployment, and vulnerability to natural disasters or terrorist acts; (2) measure their magnitude and seriousness; (3) review alternative policy interventions; (4) systematically assess the likely consequences of particular policy actions—intended and unintended, desired and unwanted; and (5) evaluate what, in fact, results from policy" (NRC, 2012b). The report also develops a framework for understanding how science is used in policy and points to areas where better knowledge could improve the utilization of science in policy making.

The utilization of science as an input to policy is a broad, complex field that this report cannot cover in detail. It raises questions and issues of global concern that scientists, policy makers, and citizens of nations around the world

will be wrestling with for years to come (Gluckman, 2014). At the same time, the responsible communication of results to policy makers and the public by researchers, and the adoption of best practices by governments in utilizing that input, are important components of scientific integrity that are closely related to other issues discussed in this report.

Recent efforts to define and implement best practices in utilizing science for policy making have focused on the development of clear policies and procedures and the utilization of transparent processes. For example, a 2009 report of the Bipartisan Policy Center explored the need for clearer policies governing the disclosure of relevant relationships by potential members of federal advisory committees, including expert testimony and consulting relationships, to prevent conflicts of interest in these activities (BPC, 2009).

As discussed in Chapter 3, the Obama administration launched an initiative in 2010 to require all federal agencies to develop and adopt scientific integrity policies (Holdren, 2010). Although an analysis by the Union of Concerned Scientists concluded that the efforts of a number of agencies fell short of what is needed to "promote and support a culture of scientific integrity," the universal adoption of such policies is certainly an important step (Grifo, 2013).

Box 9-4 provides a best practices checklist for research sponsors and users of research.

**BOX 9-4**
**Best Practices Checklist for Research Sponsors and Users of Research**

**Aligning Policies with Research Integrity**

- Maintain clear policies on research misconduct, and implement them consistently.
- Increase awareness of how policies and practices affect research integrity and quality, and act on that knowledge.
- Work to harmonize policies and practices across agencies, sectors, and national borders.

**Public Access to Data and Code**

- Develop data and code access policies for extramural grants appropriate to the research being funded, and make fulfillment of these policies a condition of future funding.
- Cover the costs borne by researchers and institutions to make data and code available.
- Practice transparency of data and code for intramural programs.
- Promote responsible sharing of data in areas such as clinical trials.
- Practice impartiality and transparency in utilizing research for the development of policy and regulations.

### Societies

According to one perspective on the role of scientific societies in fostering research integrity, "As visible, stable, and enduring institutions, scientific societies serve as the custodian for a discipline's norms and traditions, transmitting them to their members and helping to translate them into accepted research practices" (Frankel and Bird, 2003). The focus here is on disciplinary societies, although it should be noted that the largest general professional association of scientists, the American Association for the Advancement of Science, has been active over the years in a number of areas related to research integrity. Several members of the committee met with a large number of scientific society representatives as part of this study, discussing the concerns and issues facing societies and learning about what they are doing to foster integrity. Many societies publish journals as one of their core activities, and best practices associated with publishing are covered above.

Honorific academies can also play a constructive role in fostering research integrity in their national contexts, and interacademy networks can contribute at the international level by developing and disseminating guidelines and educational materials (ESF-ALLEA, 2011; IAP, 2016; NAS-NAE-IOM, 2009b).

**Best Practice S-1. Standards and Education. Serve as a focal point within their disciplines for the development and updating of standards, dissemination of best practices, and fostering RCR education appropriate to the discipline.**

The specific areas where many societies are active, apart from those related to publication, are the formulation of codes of conduct and educational efforts (Macrina, 2007). *Responsible Science* asserted that societies should play a key role in developing guidelines for research conduct appropriate to their specific fields (NAS-NAE-IOM, 1992). Many societies developed codes of conduct when research misconduct became a prominent issue in the late 1980s and 1990s, covering issues such as data handling, authorship, mentoring, and research misconduct. An American Association for the Advancement of Science survey undertaken in 2000 reported on the content and subject matter coverage of society ethics codes (Iverson et al., 2003). The American Society for Microbiology, for example, developed its first code of conduct in 1988, and it has been revised several times since (Macrina, 2007). This points to the importance of regularly updating codes of conduct in order to keep pace with changing research practices within disciplines and new ethical issues.

Societies have been active in fostering RCR education. One mechanism for doing this is through workshops or symposia held during the society's annual meeting (Iverson et al., 2003). ORI has provided support for these efforts (Macrina, 2007). Societies can also develop case studies and other educational materials that illustrate ethical issues that can arise in their disciplines. One example is

**BOX 9-5**
**Best Practices Checklist for Scientific Societies and Professional Organizations**

**Serve as a Focal Point for Developing Standards and Discipline-Specific Educational Materials**

- Serve as focal point for developing and communicating disciplinary standards to foster research integrity.
- Develop codes of conduct and keep them updated.
- Foster discipline-specific RCR education.

the American Physical Society, which developed an extensive set of case studies in the mid-2000s following several high-profile cases of research misconduct in physics (APS, 2004).

Box 9-5 provides a best practices checklist for scientific societies and professional organizations.

# 10

# Education for the Responsible Conduct of Research

***Synopsis:*** *Responsible conduct of research (RCR) education programs have become more common in recent years, partly as a result of policy changes such as the National Science Foundation's mandate that students supported by NSF research grants receive RCR education. Knowledge as to how best to provide such education is still developing. RCR educators are seeking to understand and articulate the most appropriate goals for such education, the most effective methods to be used, and the best formats in which to provide training. Improved assessments of the effects of RCR education can help develop approaches that will support both researchers and the broader institutional climate in which research takes place. RCR education can be a significant component in improving research integrity, and it will be most effective when undertaken as one of a broader set of strategies that encourage responsible conduct and discourage research misconduct and detrimental research practices. This chapter references a paper prepared for the project by Michael D. Mumford, which is included as Appendix C of this report.*

## PUTTING RCR EDUCATION ACTIVITIES IN CONTEXT

This report emphasizes that the system of research and the environments in which research is conducted should both be addressed because each strongly affects how individual researchers behave. Those who enter science and engineering learn, one way or another, about the norms and practices of the research enterprise. They are socialized into research environments and must understand something about these environments to succeed in their careers. Too often, the socialization or training they receive is often ad hoc, on the job, and not intentionally provided. Responsible conduct of research (RCR) education is important because the careers of all researchers can be significantly affected by lack of attention to responsible research practices. Thus, it is equally important in industrial, governmental, and academic settings, as well as in both private and public institutions.

One way of framing RCR education is as an *intervention* to improve the ethical conduct of investigators (see Appendix C). However, this framing can suggest that RCR education is a response to a problem and that it is external to research. An alternative, more appropriate, and likely more effective approach is to think

of RCR education as an integral part of research because RCR education aims to ensure that the knowledge, skills, and awareness essential to responsible research are intentionally, explicitly, and accurately conveyed.

Framing responsible research as the norm is both the best frame for RCR education and an essential objective of that education. RCR education targeted to individuals is designed to influence the way they understand the research enterprise and how they make decisions. In targeting individuals, the committee hopes RCR education affects attitudes and actions in ways that ultimately influence the research environment. Additionally, RCR education within an institution can create conduits for communicating and fostering a more positive institutional climate. Open discussion of ethical issues can contribute to collective openness within an institution (Anderson, 2007).

If RCR education is to be seen as more than an intervention, integration of such education into the research endeavor is key. Nominally, this includes instruction in the norms and practices of research across many and varied disciplines. Ideally, RCR education should be incorporated into the socialization and training students experience on the job, whether in the laboratory or in the myriad other locations where researchers do their work.

## FEDERAL REQUIREMENTS FOR RCR EDUCATION

The National Institutes of Health began requiring RCR education in 1989 and continues to expand and refine those requirements (NIH, 2009). In 2007 the America COMPETES Act mandated that all trainees funded by the National Science Foundation (NSF) receive RCR training as well (NSF, 2009). In 1997, NSF instituted a broader impacts criterion for the evaluation of NSF proposals, which required researchers who submit proposals to NSF to address the broad impacts of their research on society. It can be argued this was a move in the direction of expanded RCR education, but the 2007 legislation made the requirement explicit.

NSF formally established its RCR requirement in 2009, which it explained as follows:

> [E]ducation in RCR is considered essential in the preparation of future scientists and engineers. The COMPETES Act focuses public attention on the importance of the national research community's enduring commitment and broader efforts to provide RCR training as an integral part of the preparation and long-term professional development of current and future generations of scientists and engineers.

The National Institutes of Health and NSF requirements have been a major impetus for the expansion of RCR educational activities. The objectives, goals, and benefits of RCR education provide additional incentives for this expansion.

## THE OBJECTIVES, GOALS, AND BENEFITS OF RCR EDUCATION

In thinking about the aims of RCR education, it is helpful to distinguish objectives, goals, and benefits. Objectives are the broad aims of RCR education; they are what RCR education seeks to achieve in the long term and as part of a diverse set of activities. Achievement of objectives may not be measurable within a particular course or activity. For example, it may not be possible to determine whether or to what extent a particular course or course module has contributed to the objective of reducing the incidence of research misconduct.

In contrast with objectives, goals are narrower in scope and more specific. Goals might be measured in the assessment of a particular activity. For example, a goal might be to ensure that researchers are aware of codes of conduct. Goals are related to objectives in the sense that they may be adopted because of their contribution to a broader objective. For example, a course might have the goal of improving ethical decision making. This goal in turn contributes to the broader objective to ensure the integrity of research.

In addition to objectives and goals, RCR education may provide benefits not identified as an objective or a goal. For example, improving the retention of researchers who might otherwise have left the field due to disappointment in the practiced norms of research is an advantage of RCR education, but may not be a specified objective or goal. Objectives, goals, and advantages overlap and are not always easy to distinguish.

### Overall Objectives

Among the major objectives identified in the literature on RCR education are the following:

- Ensuring and improving the integrity of research; promoting good behavior and quality research conduct;
- Preventing bad behavior; decreasing research misconduct;
- Making trainees aware of the expectations about research conduct within the research enterprise and as articulated in various federal, state, institutional, and professional laws, policies, and practices that exist;
- Making practitioners and trainees aware of the uncertainty of some norms and standards in research practices due to such factors as changes in the technology used in research and the globalization of research;
- Promoting and achieving public trust in science and engineering;
- Managing the impact of research on the world beyond the lab, including society and the environment.

### Goals for Educational Activities

These broad objectives have been formulated into more concrete goals to be achieved by particular forms of RCR education. For example, Michael Davis (Davis and Feinerman, 2010) identifies four goals for RCR education: ethical sensitivity (being able to recognize ethical issues), ethical knowledge, ethical judgment, and ethical commitment. The report of a 2008 workshop organized by the National Academy of Engineering and sponsored by NSF describes a set of skills to be developed in RCR education as follows (NAE, 2009):

- Recognizing and defining ethical issues;
- Identifying relevant stakeholders and sociotechnical systems;
- Collecting relevant data about the stakeholders and systems;
- Understanding stakeholder perspectives;
- Identifying value conflicts;
- Constructing viable alternative courses of action or solutions and identifying constraints;
- Assessing alternatives in terms of consequences, public defensibility, and institutional barriers;
- Engaging in reasoned dialogue or negotiations;
- Revising options, plans, or actions.

Each of the skills listed is an activity that contributes to ethical decision making. For example, to behave responsibly one has to recognize that a situation poses an ethical problem; then, if one is to act in the situation, it is important to identify the relevant stakeholders, construct and assess alternative courses of action, and so on.

For many RCR educators, decision making, and specifically ethical decision making, should be the primary focus of RCR education. Kalichman has succinctly argued for this; he identifies three possible objectives for RCR education: decreased research misconduct, increased responsible conduct of research, and improvements in ethical decision making (Kalichman, 2012). He rejects the first two as unknowable and focuses on ethical decision making.

As can be seen from the preceding, the challenge of RCR education derives in part from the nature of ethics teaching, which involves conveying knowledge, developing skills, shaping attitudes, and affecting behavior.

## THE EFFECTIVENESS OF RCR EDUCATION AND ITS ASSESSMENT

As a formal undertaking, RCR education is still in the early stages of its development. Experience over the last several decades has provided some basis for going forward, but the state of knowledge in the field is far from mature.

A particular focus within RCR education has been the assessment of its effects, but assessment of education in ethics is a relatively new field. The chal-

lenges of assessing ethics education are intertwined with the challenges of identifying and specifying the objectives and goals of such education.

In his review of assessments of the effectiveness of RCR training, Mumford concludes that the evidence indicates weak but positive effects (see Appendix C). Restricting his evaluation to improvements in ethical decision making, Kalichman characterizes the evidence as equivocal at best (Kalichman, 2012).

In the interaction between assessment, objectives, and goals, a "chicken-and-egg" problem may arise. One of the purposes of assessment is to help identify the most effective approaches to take in RCR education. At the same time, useful assessment depends on identifying and articulating measurable objectives and goals. In other words, specifying appropriate objectives and goals for RCR education is critical to assessment, yet assessment informs the selection of appropriate objectives and goals. Hence, a point of caution is appropriate here. In the interplay between assessment and RCR education, assessment should follow, not lead. One standard criticism of assessment is captured in the phrase "measures become targets." The concern is that assessment will take forms or produce results that have too strong an influence on the structure or content of RCR education. RCR educators may teach or train to the assessment tool rather than continue to reflect on what are the important objectives, measurable or not.

One of the challenges in identifying what constitutes strong RCR education programs is the varied approach to assessment of RCR education, which in part arises from diverse perspectives on the educational goals of RCR courses. Achievement of the broad objectives and ancillary benefits described earlier can be even more difficult to assess, since many accrue over the long term and require large populations to demonstrate statistical significance (such as effects on the incidence of misconduct or rates of retention in science).

Despite the challenges of assessment, the Project for Scholarly Integrity at the Council of Graduate Schools and discussion at the Ethics Education in Science and Engineering Workshop at the National Academy of Engineering both support the assertion that assessment is critical to successful and sustainable RCR programs (CGS, 2012; NAE, 2009). Kalichman describes four key goals of RCR education that could be assessed: (1) increases in knowledge of issues and practices, (2) increases in skills related to ethical decision making and conflict management, (3) improved attitudes toward open communication and respect of issues, and (4) improvements in behavior and choices (Kalichman, 2012). Mumford describes key goals of RCR education as improvements in ethical decision making, perceptions of ethical climate, and knowledge (Appendix C). Mumford goes on to describe assessment measures of RCR education as measures of performance (such as decision making in ethics cases), knowledge (such as the results of an exam on human subjects regulation), climate (such as the extent to which individuals endorse ethical behaviors), products (such as self-reflection exercises), or organizational outcomes (such as a drop in the incidence of ethical violations) (Appendix C). Most assessment efforts for RCR education

have focused on improvements in ethical decision making (Antes et al., 2010; Bebeau, 2002; Mumford et al., 2008; Pimple, 2001; Schmaling and Blume, 2009) and/or knowledge (Elliott and Stern, 1996; Pimple, 2001; Schmaling and Blume, 2009). While more difficult to assess, some have attempted to assess behavioral choices (Anderson et al., 2007a; Wester et al., 2008).

## STRATEGIES AND FORMS OF RCR EDUCATION

How RCR education is assessed depends on the goals of the educational activity, and the goals in turn depend on the form the education takes. For example, a 1-hour module on data sharing or conflicts of interest will have narrower and different goals than would a full-semester course or a guest lecture series.

A number of formats have been adopted to provide RCR training for science and engineering trainees, including graduate students, postdoctoral trainees, and undergraduate students. These include stand-alone courses (DuBois et al., 2008; Elliott and Stern, 1996; Kalichman and Plemmons, 2007; Plemmons et al., 2006; Powell et al., 2007; Schmaling and Blume, 2009), seminar/workshop series that are either concentrated within a short period (such as an ethics week) or spread across a longer term (Antes et al., 2009; Clarkeburn et al., 2002; Ferrer-Negron et al., 2009; Fischer and Zigmond, 2001), ethics across the curriculum approaches that embed ethics materials into science and engineering coursework (Antes et al., 2009; Canary and Herkert, 2012; Davis and Riley, 2008; Frugoli, 2002; Smith et al., 2007), web-based training modules (Braunschweiger and Goodman, 2007; DuBois et al., 2008; Sieber, 2005), hybrid programs that use combinations of these approaches (Canary and Herkert, 2012), and laboratory-based interactions (Canary and Herkert, 2012).

Studies of the relative efficacy of these different approaches remain limited, but some modest positive results have been found for most approaches (see Appendix C; see also Antes et al., 2009; Elliott and Stern, 1996; Ferrer-Negron et al., 2009). However, one study has also found some negative effects of RCR education, particularly when students internalize the ideas that decisions regarding ethical issues have the potential to derail careers and that other researchers are unethical (Antes et al., 2010). Another found negative effects of RCR training but positive effects with RCR mentorship (Anderson et al., 2007a). Others have found a lack of change in assessment measures (Kalichman and Friedman, 1992).

Antes has argued that separate courses and seminars are more successful in ethical decision making than embedded programs (Antes et al., 2009). Others argue that RCR education content embedded in disciplinary or methods courses can also be successful (Davis and Riley, 2008). Web-based training has received the most criticism for possible lack of effectiveness, particularly when it is a pass/fail endeavor with little interpersonal interaction (NAE, 2009). Teaching RCR in a purely online format raises issues that surround online education more generally,

such as the difficulty of interacting with an instructor personally. In addition, the goals of teaching ethics, especially the skill, attitude, and behavior components, may not be as amenable to an online format as other material. Online formats tend to focus on knowledge and can be limited in their ability to teach ethical decision making and conflict management skills, which many identify as critical components of RCR education (Appendix C; see also Kalichman and Plemmons, 2007). However, the effectiveness of such web-based training, and indeed any training approach, appears to depend on what is included and how exactly students are engaged in the materials (Antes et al., 2009). Mumford (Appendix C) and Antes et al. (2009) both point to RCR programs that involve active and cooperative formats as being more effective in developing ethical decision-making processes, and these formats can be difficult to achieve online. Canary and Herkert (2012) found the strongest efficacy in RCR programs that take a hybrid approach.

While these studies are interesting, solid research on the efficacy of these different delivery methods is scarce for several reasons, including the lack of a standard approach to assessment, a lack of agreement on the goals of RCR education, and the challenges of conducting such educational research. In addition, RCR education is profoundly affected by the context in which that education exists (NAE, 2009). As such, efficacy may need to be studied within the context of the institution and the research field.

Within these RCR curricular approaches, educational methods can also vary widely. These include lecture, discussion of professional codes, expert panels, case-based discussions (Antes et al., 2009; Bebeau, 1995; DuBois et al., 2008), presentation and discussion of moral exemplars (Harris, 2008), role playing (Brummel et al., 2010; Seiler et al., 2011; Strohmetz, 1992), ethics issues embedded in science and engineering problems, and service learning (Fitch, 2004; Pritchard, 2000). Methods that encourage interaction have generally been found to be more successful (Antes et al., 2009; NAE, 2009). However, the relative effectiveness of these methods has yet to be examined fully, and such an examination may again be limited by the lack of standardization of RCR goals and assessment methods.

Beyond improved moral reasoning, the topics in research practice that should be covered in an RCR educational program have received only limited discussion. Such topics can include issues of credit in authorship and intellectual property (including issues of plagiarism), appropriate treatment of human and/or animal subjects, issues of conflict of interest, appropriate data management (including issues of fabrication and falsification), issues in the peer review process, mentoring and employment relationships, and societal impacts of research. In addition, specific disciplines can have discipline-specific topics such as the relationship between research and clinical practice in medicine or design and manufacturing issues in engineering. Conflict management techniques and processes (including both interpersonal communication skills and knowledge and understanding of organizational and institutional dynamics and structures) have been suggested

as an important component of RCR education (Gunsalus, 1998a; Kalichman and Plemmons, 2007). Topics in RCR education also are not static. For example, emerging issues include the use and accessibility of computer code in research and the proper application of statistical methods to large datasets.

The topics discussed in this report could provide guidance on topics appropriate for inclusion in RCR education. In particular, reducing detrimental research practices and improving best practices benefit from open and active discussion among scientists. Awareness of environmental effects on individual choices could support stronger individual decision making.

## Resources for RCR Education

Since the publication of *Responsible Science* in 1992, a number of resources for RCR education have been created. Despite the inherent challenges with online education, many institutions have begun to use online training resources. The most used in the United States currently is the Collaborative Institutional Training Initiative program. This program has provided web-based training to thousands of institutions in 40 countries since 2000. Another example is the online content provided by Epigeum, a British education provider that offers courses on research ethics. A number of institutions have collected a range of online resources, and the Office of Research Integrity has produced an online video to address some elements of content.

Additional online repositories and collaborative environments for RCR educational materials include the website of the National Academy of Engineering's Online Ethics Center for Science and Engineering (www.onlineethics.org), the NSF-funded National Ethics Center (Ethics CORE) (nationalethicscenter.org), the Resources for Research Ethics Education website (research-ethics.net) sponsored by the University of California, San Diego, and the Committee on Publication Ethics' eLearning course on publication ethics for editors and publishers (COPE, 2017). Journals in the field include *Science and Engineering Ethics* (http://link.springer.com/journal/11948), which regularly publishes articles on research ethics and the teaching of research ethics, and *Research Ethics* (http://journals.sagepub.com/home/rea), which is sponsored by the Association for Research Ethics and is devoted to ethical research in human beings. Publishers have also developed RCR education materials; the majority of these materials specifically address publishing ethics. Wiley has recently released a second edition of *Best Practice Guidelines on Publication Ethics: A Publisher's Perspective* (Graf et al., 2014), Elsevier provides a Publishing Ethics Resource Kit (https://www.elsevier.com/editors/perk), and BioMed Central provides resources on the "Publication Ethics" page of its website (https://www.biomedcentral.com/getpublished/writing-resources/publication-ethics).

In 2009, the National Academies published the third edition of its widely used reference, *On Being a Scientist: A Guide to Responsible Research Conduct*,

which is intended to supplement research ethics lessons provided by institutions, research mentors, and supervisors (NAS-NAE-IOM, 2009b). Among other things, the guide discusses treatment of data, research misconduct, authorship credit, and conflicts of interest. Hundreds of thousands of print and electronic copies of the guide have been distributed since the first edition was released in 1988.

In 2016 the InterAcademy Partnership released *Doing Global Science: A Guide to Responsible Conduct in the Global Research Enterprise,* intended for use in education and training contexts on a global basis (IAP, 2016). In addition, other textbooks, compilations of case studies, and other written materials are available (e.g., Penslar, 1995).

## RCR EDUCATION IN THE BROADER CONTEXT

As mentioned before, RCR training is most effective when it is one element in a comprehensive approach to improve an institution's system of research. If a comprehensive approach is not taken, aspects of the broader system may undermine the effectiveness of RCR education. For example, when faculty members and administrators or managers express a lack of enthusiasm or even disdain for RCR training, students and postdoctoral researchers get the message and may come to see RCR training as a regulatory burden or even believe that responsible conduct is not important to their research or careers. Research environments can convey this message in a number of subtle and unintended ways. For example, having RCR materials presented in the classroom only as the last lecture in a full-semester course or as a guest lecture on a day the professor will be absent may be interpreted as reflecting the professor's lack of interest in the material.

Limitations on instructional time and the demands of research can limit the amount of RCR education that can be provided. For example, institutional pressures on research productivity have been shown to have negative effects on responsible conduct, and such pressures can also affect the willingness of research mentors to allow time for RCR education (Anderson et al., 2007b).

Although all the participants involved in research are important in creating an environment that is conducive to the responsible conduct of research, two categories of participants are especially important—institutional leaders and mentors. As Chapter 6 describes, institutional leadership and climate can be either a support or a barrier to effective RCR education. The Council of Graduate Schools' project on scholarly integrity has recommended engaging the leadership of institutions as a critical part of any sustainable and effective RCR program. Beyond supporting RCR educational programs within the institution, institutions should be looking more broadly at educational and other activities that encourage research integrity. For example, in Reason's 2000 *BMJ* article, a suggestion was made for institutions to move from blame-and-shame methods for dealing with misconduct to reporting and feedback. Such a reporting-and-feedback dynamic

might be created and nurtured through the RCR educational process as an element of a broader institutional program.

Mentors have particularly important roles because they advise aspiring researchers and because young researchers look to them as role models. The words and actions of a research mentor can both positively and negatively impact ethical behaviors and potentially support or undermine RCR educational efforts (Anderson et al., 2007a; Antes et al., 2010; Wright et al., 2008). Thus, to have an effective RCR environment, mentors must understand the effects of their behavior on young researchers and must be held accountable for conveying the importance of responsible conduct to their trainees. Ideally, mentors, other research scientists, and institutional leaders all actively participate in RCR discussions, since everyone involved in such discussions can benefit from open and honest discourse regarding best practices and detrimental research practices.

## A COMPREHENSIVE APPROACH TO RCR EDUCATION

RCR education cannot be considered a total solution to the problem of ensuring responsible conduct. Rather, it should be seen as one component in a comprehensive approach that includes improving mentorship and institutional climate.

RCR education can continue to develop through the identification of a strong set of educational goals, the development of new educational tools, and the refinement of assessments. It also can be expanded to include not just trainees but research mentors, principal investigators, and institutional leaders in discussions of research ethics. Such involvement will contribute to a positive institutional climate and a greater collective openness.

In particular, since research mentors are so influential in the development of ethical behavior, RCR educational efforts should examine ways to use this relationship more productively to foster responsible conduct.

# 11

# Findings and Recommendations

The committee's findings and recommendations reported in this chapter are based on its examinations of changes in the research environment since the 1992 *Responsible Science* report and on the committee's consensus on the means by which the U.S. research enterprise and its participants might best foster scientific integrity in the changing environment (NAS-NAE-IOM, 1992). Despite the intensification and acceleration of forces originally discussed in *Responsible Science,* and the emergence of some trends that were not apparent then, the core values of the responsible conduct of science have not changed and should not change. These core values include objectivity, honesty, openness, accountability, fairness, and effective stewardship. The committee has structured its recommendations around these values and an understanding that research is conducted as part of a larger social enterprise. The resources produced by the research enterprise—including knowledge and highly trained people—are intended to benefit the public. Scientists are provided with opportunities and freedom to pursue new knowledge and train future scientists with the implicit understanding that they are responsible for the conduct of their research and the reliability of the knowledge they produce and that they must conduct their research responsibly as a duty to the public.

## UNDERSTANDING THE ISSUES

### Changing Environment

A number of changes in the research environment that were identified in the early 1990s as problematic for maintaining principles of research integrity and good scientific practices have generally continued along their long-term trend

lines. These include growth in the size and scope of the research enterprise, the increasing need for and complexity of collaboration, the expansion of regulatory requirements, and an increased focus on industry-sponsored research.

Several important new trends that were not examined in *Responsible Science* have also emerged, including the pervasive and growing importance of information technology in research, the globalization of research, and the increasing relevance of knowledge generated in certain fields to policy issues and political debates. These changes—the growing importance of information technology in particular—have led to important shifts in the institutions that support and underlie the research enterprise, such as science, engineering, technology, and medical publishing. The understanding of how colleagues, incentives, and environments influence ethical decision making has also advanced significantly. New challenges must be straightforwardly addressed to support researchers, research institutions, journals, and sponsors in their efforts to foster integrity, prevent and discourage research misconduct and detrimental research practices, and respond to these problems when they occur.

### Updating Concepts: The Role of Detrimental Research Practices

Much of the discussion, thinking, and actions aimed at fostering research integrity has revolved around the actions of miscreant individuals in committing acts of research misconduct and its components—fabrication, falsification, and plagiarism. Actions that *Responsible Science* characterized as questionable research practices have received less attention. The accumulation of knowledge has brought the critical need to address these elements to the fore. Actions such as failing to retain or share data and code supporting published work in accordance with disciplinary standards, practices such as honorary or ghost authorship, and using inappropriate statistical or other methods of measurement and data presentation to enhance the significance of research findings are clearly detrimental to the research process and may impose comparable or even greater costs on the research enterprise than those arising from research misconduct. The committee believes that identifying these actions as detrimental research practices (DRPs) will be helpful in focusing attention and developing approaches to discourage and minimize them.

At the same time, based on better insight and understanding of the importance of environmental influences on individual choices, the environments in which research is performed need to be thoughtfully assessed and shaped. In addition to DRPs committed by individual researchers, organizations such as research institutions, research sponsors, and journals may also take actions that constitute detrimental research practices, often by failing to acknowledge or act upon implicit or explicit incentives and reward systems that can undermine the integrity of the research enterprise.

## Incidence and Costs

The incidence of discovered research misconduct is tracked by official statistics, survey results, and analysis of retractions, and all of these indicators have shown increases in recent years. However, it is difficult to estimate precisely the incidence of misconduct in relation to an established baseline and to determine trends. It is possible to say that while research misconduct is unusual, it is not rare. High-profile cases continue to appear regularly from around the world at the same time that the overall size of the research enterprise has vastly expanded. A variety of DRPs appear to be unfortunately fairly common, at least in the fields and disciplines that have been studied. Examining specific cases of misconduct shows that tolerance for DRPs enables misconduct and leads to delays in uncovering it.

Both research misconduct and DRPs impose significant costs on the research enterprise, including careers that are destroyed or sidetracked, the financial costs to taxpayers of fraudulent or otherwise irreproducible research and work done to extend it, reputational costs to institutions, and the costs of investigations. Particular cases of misconduct and DRPs have also negatively affected society at large, such as a purported finding of a causal link between a widely used childhood vaccine and autism that has played a role in discouraging vaccinations. Such cases cause direct harm and also damage societal trust in the research endeavor. Beyond questions of needless human suffering, the total scale of monetary costs from research misconduct and detrimental research practices may run from several hundred million dollars up to multiple billions of dollars per year in the United States alone.[1]

Some of the actions and approaches needed to foster integrity recommended below do not have major costs associated with them. Others do, whether at the lab, institutional, or disciplinary level, and these costs are difficult to estimate. The research enterprise and sponsors may need to confront the need to spend more per research output, with the end result being fewer or slower research outputs. But those outputs—and the research cultures and environments in which they arise—will be much more robust, especially in those disciplines that have seen major issues of lack of robustness and trust.

## Understanding the Causes

Why people engage in criminal or other deviant behavior and the conditions that encourage or discourage such behavior are issues of perennial interest in the behavioral and social sciences. Recent work has contributed useful insights that

[1] Chapter 5 contains a detailed discussion of the current state of knowledge concerning the various costs and consequences of research misconduct and DRPs and how their scale and scope might be estimated.

are relevant to understanding why and under what conditions researchers commit misconduct and engage in DRPs.

Some past assumptions and assertions have held that the character of scientists as searchers for truth and strong traditions of mentorship would limit those who would commit research misconduct to a few "bad apples." However, evidence from recent years makes it clear that scientists are not immune to the environmental forces that contribute to deviant behavior in all professions, nor are they exempt from a variety of cognitive biases that are a normal part of the human condition (Mazar and Ariely, 2015). The environments in which researchers are educated, socialized, and perform their work require significant attention.

Current patterns of U.S. research funding and organization have been identified by some leading scientists as creating hypercompetitive research environments that are damaging the long-term health of research in some of the largest fields and disciplines (Alberts et al., 2014). These hypercompetitive environments contain characteristics that behavioral and social sciences research suggests facilitate and encourage detrimental and deviant behavior. While addressing larger structural issues in U.S. research funding and organization is beyond the scope of this study, more research on the causes of research misconduct and detrimental research practices is needed to develop better strategies for prevention, as well as specific steps to assess the integrity of research environments and to act on the findings of such assessments to implement good practices and foster sound research environments.

### The Need for More Robust Approaches

Research misconduct and DRPs need to be addressed in several ways. The primary means include (1) efforts to prevent them through responsible conduct of research (RCR) education and environmental assessment and improvement; (2) efforts to uncover research misconduct, investigate, and take corrective actions through the efforts of researchers, institutions, federal and private research sponsors, and journals; and (3) efforts to discourage and eliminate DRPs through the implementation of standards and best practices, such as effective mentoring, requirements for data and code sharing, and implementation of greater transparency in reporting results.

The committee examined the status of efforts in all three areas and concluded that improvements are needed across the board. The findings and recommendations that follow provide a roadmap for the actions that need to be taken.

## FINDINGS AND RECOMMENDATIONS

The committee reaffirms the central recommendation from *Responsible Science* that formally places the primary responsibility for acting to define and strengthen basic principles and practices for the responsible conduct of research

on individual scientists and research institutions. At the same time, the committee based its recommendations on its understanding that the integrity of research depends on creating and maintaining a system and environment of research in which institutional arrangements, practices, policies, and incentive structures support responsible conduct. Fostering research integrity is an obligation shared not only by individual researchers but also by leaders and those involved with all organizations sponsoring, conducting, or disseminating research, including corporate and government research organizations.

The committee also endorses the definition of research misconduct recommended in *Responsible Science* while recommending refinements in its use. In particular, through its examination of current practices advancing research integrity and responses to deviations, the committee became aware of variations in federal approaches to the evaluation of plagiarism that need to be harmonized. The following findings and recommendations are intended to serve as a framework for actions that will improve knowledge of research misconduct, detrimental research practices, and contributing factors; strengthen approaches to addressing them; and ultimately lead to a significant reduction or even elimination of these behaviors and the risks that they pose to the research enterprise.

***FINDING A: Developing and implementing improved approaches to fostering research integrity and meeting the current threats to integrity posed by research misconduct and detrimental research practices are urgent tasks. These improved approaches should reflect an understanding of the complex interactions among the many components of the research enterprise and its multiple stakeholders.***

The research enterprise is a large, diverse, and complex system, and society invests considerable resources in it. Research misconduct and detrimental research practices constitute long-term threats to the research enterprise's ability to deliver the benefits expected by society. While the values and ideals of science should remain unchanged, the experience of the past several decades drives home the lesson that significant changes in the practices and institutional arrangements of the research enterprise are necessary to strengthen the self-correcting mechanisms of science. Developing and implementing these practices and approaches will require us to better understand how research environments and the incentives created by structural relationships among the institutions of science can support or undermine the efforts of individual researchers to behave responsibly.

The first set of recommendations targets the broad, long-term need for sustained, cooperative efforts by the major components of the research enterprise: individual researchers; research institutions; research sponsors; science, engineering, technology, and medical journal and book publishers; and scientific societies. They also cover the shorter-term need to improve research environments by assessing and then addressing identified weaknesses.

**RECOMMENDATION ONE: In order to better align the realities of research with its values and ideals, all stakeholders in the research enterprise—researchers, research institutions, research sponsors, journals, and societies—should significantly improve and update their practices and policies to respond to the threats to research integrity identified in this report.**

Lack of attention to or tolerance of detrimental research practices by stakeholders makes it more difficult to expose misconduct, wastes human and financial resources, impairs the overall quality of research, and diminishes public trust in science. In addition, weaknesses in the system for identifying, investigating, and sanctioning research misconduct—most notably unevenness in the policies and capabilities of research institutions and journals—create barriers to uncovering misconduct and taking corrective action. Changes in the funding and organization of research are affecting institutional and laboratory environments in ways that can undermine incentives to behave responsibly. For example, Alberts et al. (2014) noted,

> As competition for jobs and promotions increases, the inflated value given to publishing in a small number of so-called 'high impact' journals has put pressure on authors to rush into print, cut corners, exaggerate their findings, and overstate the significance of their work. Such publication practices, abetted by the hypercompetitive grant system and job market, are changing the atmosphere in many laboratories in disturbing ways.

Similarly, in industrial R&D organizations, pressures associated with regulatory approvals or commercial release may create disincentives for full data transparency or biases that promote conclusions of safety and efficacy. Finally, changes in the research environment such as technological advances and globalization are making it more difficult and complex for all stakeholders in the enterprise to update and ensure adherence to best practices.

The checklists presented in Chapter 9 should form the basis of strategies to refine and implement best practices by researchers, research institutions, research sponsors, journals, and societies.

**RECOMMENDATION TWO: Since research institutions play a central role in fostering research integrity and addressing current threats, they should maintain the highest standards for research conduct, going beyond simple compliance with federal regulations in undertaking research misconduct investigations and in other areas.**

In order to maintain the highest standards for research conduct, research institutions need to exercise vigilance in several distinct areas:

- Creating and sustaining a research culture that fosters integrity and encourages adherence to best practices through effective education and training and other mechanisms;
- Monitoring the integrity of research environments through internal assessments and multi-institution benchmarking exercises, and acting on the results;
- Ensuring that institutional policies and processes to investigate and address allegations of research misconduct are robust and generate just and timely outcomes; and
- Ensuring that senior institutional leaders such as the president, other senior executives, administrators, and faculty leaders are guiding and actively engaged in these efforts.

Because they are the facilitators and stewards of research activity, as well as the employers and educators of researchers, research institutions (including academic/nonprofit, industrial, and governmental organizations) will play a central role in determining how well the research enterprise as a whole fosters research integrity and addresses current threats. Institutions can undertake this important work collaboratively—through related organizations and associations—as well as in partnership with other stakeholders, such as federal and private research sponsors and science, engineering, technology, and medical journal and book publishers.

The key responsibilities for research institutions fall into four areas. The first is creating and sustaining a research culture that fosters integrity and encourages adherence to best practices. The leadership of universities and other research institutions, including presidents, other senior executives and administrators, and faculty leaders, has a central role to play in building and sustaining environments that promote responsible research conduct. This is not only important for fostering integrity in the research process but also will encourage science of the highest quality. This includes maintaining education and training efforts that support a culture of integrity, consistent with the current state of knowledge. Recommendation Ten describes in more detail how responsible conduct of research education and training programs should be developed and implemented.

A second task is monitoring the integrity of research environments. Such monitoring is critical to further advance understanding of how institutional structure, context, and incentives interact to buttress or detract from research integrity. The 2002 report *Integrity in Scientific Research: Creating an Environment That Promotes Responsible Conduct* explains that research organizations "engage in activities that help establish an internal climate and organizational culture that are either supportive of or ambivalent toward the responsible conduct of research" (IOM-NRC, 2002). Institutional assessment and benchmarking exercises can be important tools helping institutional leadership to fulfill this role; tools now exist that can be used to perform these assessments (Martinson et al., 2013). Research

institutions should regularly assess their research integrity climate and share information among departments and peer institutions, using approaches such as those used in the Project on Scholarly Integrity undertaken by the Council of Graduate Schools (CGS, 2012). Related organizations and associations (e.g., the Association of American Universities, Association of Public and Land-Grant Universities, Council of Graduate Schools) should contribute to this effort. Part of this activity should be devoted to further advance understanding of how institutional structures, context, and incentives interact to buttress or detract from research integrity.

Research institutions also have an obligation to implement improvements to their research environments based on the knowledge gained in these assessments. Recent efforts involving assessment of research integrity climates and sharing of information among departments and peer institutions have yielded important insights. Where institution-wide assessments identify units with particularly strong integrity environments, they should be examined and their practices should be disseminated and emulated.

The third institutional responsibility is ensuring that research institutions sustain the capacity needed to effectively investigate and address allegations of research misconduct. No institution can be expected to prevent all lapses in research integrity, but all should ensure that when problems in the conduct of research are alleged or identified, there is a prompt, effective, and documented response to the allegation. Currently, there is limited evidence to assess how institutions are performing, including specific cases that have been reported in the media, the results of surveys of institutional officials undertaken by the Office of Research Integrity, and the presentations made to the committee by federal agency officials. This evidence is discussed in Chapter 7 and indicates that, while some institutions are performing at an outstanding level in this area, others are not. This report describes several highly publicized cases from recent years in which institutional responses to alleged research misconduct or to credible questions about reported results were deficient. Given the critical role that institutions play in fostering research integrity, substantial damage can be done in cases where they fall short. The best practices for research institutions described in Chapter 9 comprise a starting point for institutional efforts.

The committee appreciates that sustaining appropriate institutional capacity can be challenging. Because of the relative infrequency of inquiries and investigations responding to allegations of research misconduct, particularly at smaller institutions, it may be difficult to maintain institutional memory in some areas. The specific examples examined by the committee also show that it is often difficult for organizations to work impartially when powerful individuals have been accused of misconduct or when the institution's own financial or reputational interests are involved. The Research Integrity Advisory Board proposed below can serve as a resource as institutions seek to maintain the highest standards in how they address lapses in integrity.

Institutions are organized differently, so they should have the flexibility to develop and implement their policies in ways that make sense for them. In some institutions, formal responsibility for responding to research misconduct allegations lies with the graduate dean or vice president for research. In others, a designated research integrity officer or compliance officer who reports to the vice president for research might have this responsibility. In all cases, this official—whose responsibilities might also include oversight of education, training, and assessment as discussed below—should have direct access to the president and other institutional leaders.

One insight from specific cases is that the existence of multiple channels for raising concerns even prior to making allegations can be very helpful. For example, some institutions have had success with an ombuds system independent of those formally responsible for responding to allegations.

A fourth responsibility is ensuring that senior institutional leaders are guiding and actively engaged in the preceding three tasks. For example, when institutional leaders are accessible and knowledgeable about institutional capacity to address allegations of misconduct, they are in a position to be helpful in keeping people and processes on track when specific allegations arise. Should later events call into question the rigor of an institutional response to allegations of misconduct in research, top institutional leadership should be expected, as a matter of course, to examine the shortcomings of the process and share lessons learned with the larger community of scholars as a contribution to improvement of research integrity across the community. Institutional leaders must be accountable for the quality of responses to questions about research integrity.

Institutional leaders are also in the best position to implement changes based on the results of research integrity climate assessments, and they can communicate directly and regularly institutional standards and expectations as well as the importance of the quality of research conducted under institutional auspices.

**RECOMMENDATION THREE: Research institutions and federal agencies should work to ensure that good-faith whistleblowers are protected and that their concerns are assessed and addressed in a fair, thorough, and timely manner.**

Those who raise concerns about the integrity of research, often referred to as whistleblowers, can play a critical role in supporting best practices in research and in uncovering research misconduct, as described in Chapter 7. Individuals closest to the research are in the best position to identify and correct problems as early as possible and can be expected to play this role for the foreseeable future. Inadequate responses to expressed concerns have constituted a critical point of failure in many cases of misconduct where investigations were delayed or sidetracked. Those who raise concerns are typically the most vulnerable participants in the system, holding little institutional power or status. Research institutions

and federal agencies should understand the implicit bias that exists against those who in good faith raise fact-based concerns about the integrity of research. The report discusses several alternative approaches to strengthening whistleblower protections that should be considered and implemented.

**RECOMMENDATION FOUR: To provide a continuing organizational focus for fostering research integrity that cuts across disciplines and sectors, a Research Integrity Advisory Board (RIAB) should be established as an independent nonprofit organization. The RIAB will work with all stakeholders in the research enterprise—researchers, research institutions, research sponsors and regulators, journals, and scientific societies—to share expertise and approaches for addressing and minimizing research misconduct and detrimental research practices. The RIAB will also foster research integrity by stimulating efforts to assess research environments and to improve practices and standards.**

While various groups, institutions, and individuals are doing valuable work to foster and promote research integrity in the United States, no permanent organizational focus for efforts to foster research integrity at a national level currently exists. The RIAB would provide a continuing organizational focus for fostering research integrity that cuts across disciplines and sectors. It should be established independent of government.

The RIAB would perform several functions, including:

- Working with public and private research sponsors to develop improved practices and approaches to addressing research misconduct and fostering integrity. For example, the RIAB could serve as a forum for the discussion of issues where community consensus currently does not exist (such as what the appropriate penalties for research misconduct should be) or where current disparate approaches should be harmonized (such as the implementation of the federal research misconduct policy in areas such as plagiarism).
- Working with science, engineering, technology, and medical journal and book publishers to develop improved practices and approaches. The biannual Journal Summit organized by the National Academy of Sciences generates a number of useful ideas that could be explored further by the RIAB.
- Identifying important topics and questions related to research misconduct and research integrity, including pathways to improve research environments and RCR education, where research could produce valuable insights, and perhaps serve as a mechanism for commissioning such research.

- Working with research institutions, institutional officials, and groups such as the new Association of Research Integrity Officers to identify and develop resources aimed at improving institutional capability to respond to research misconduct allegations and sustain environments that encourage responsible conduct. These resources could include just-in-time training materials, referrals to experts with relevant scientific and/or legal knowledge who could be consulted on specific cases, and help with organizing external review of investigation committee task statements and reports.

The RIAB will have no direct role in investigations, regulation, or accreditation. Rather, it will serve as a neutral resource based in the research enterprise that helps the research enterprise foster integrity in a changing environment. It will work best as an independent, nonprofit organization with a small permanent staff of three or four people, supplemented by fellows and consultants. An annual budget of about $3 million would be adequate. The RIAB would be governed by its members, with a rotating executive committee selected to develop strategy and oversee operations. Funding would come in the form of regular contributions from members such as the major public and private sponsors of research, universities and other research institutions, industrial members, scientific societies, and science, engineering, technology, and medical journal and book publishers. Further discussion of the RIAB and consideration of alternatives is contained in Chapter 8.

***FINDING B: Ensuring greater openness and accountability in science is essential to fostering research integrity and improving research quality. Establishing and agreeing on new standards and building the infrastructure needed to implement those standards will require collaborative, focused efforts on the part of the research enterprise and its stakeholders.***

The values of openness and accountability make transparency and striving for reproducibility of scientific findings central to the responsible conduct and dissemination of research. As technological advances and other shifts continue to transform scientific work, responsive changes to standards and practices are needed. Examples from recent years show that some cases of fabrication and falsification have been uncovered relatively quickly by researchers seeking to replicate the work when data were available. In other examples, failure to require that researchers provide access to data and code has been associated with delays in uncovering lapses in integrity. Clarifying and updating authorship standards, implementing data- and code-sharing requirements, securing adherence to existing requirements, and heightened attention to appropriate use of sound statistical methods will help to foster integrity by facilitating the processes by which research results are confirmed or refuted.

Facilitating broader access to data and code can also help to accelerate the advance of knowledge. In recent years, the problem of irreproducibility of research results has attracted increasing attention and concern. It is important to note that some research results will not be reproducible even where there are no mistakes or lapses in integrity. Researchers can ensure openness, honesty, accountability, and transparency, but cannot completely ensure the reproducibility of their work. The baseline responsibility of researchers, institutions, journals, and sponsors is to ensure that published research provides enough information about methods and tools that other researchers attempting to replicate the work could succeed or, if not, could provide compelling evidence that the work could not be reproduced.

All of the actions outlined in this set of recommendations are aimed at ensuring higher levels of openness and accountability, which are essential to strengthening the operation of the scientific process and for ameliorating many of the weaknesses that are apparent in current systems and practices. As pointed out in various parts of the report, several of these areas of weakness have seen positive movement in recent years, but efforts on the part of one or more research enterprise stakeholders could bring practices into better alignment with scientific ideals.

**RECOMMENDATION FIVE: Societies and journals should develop clear disciplinary authorship standards. Standards should be based on the principle that those who have made a significant intellectual contribution are authors. Significant intellectual contributions can be made in the design or conceptualization of a study, the conduct of research, the analysis or interpretation of data, or the drafting or revising of a manuscript for intellectual content. Those who engage in these activities should be designated as authors of the reported work, and all authors should approve the final manuscript. In addition to specifying all authors, standards should (1) provide for the identification of one or more authors who assume responsibility for the entire work, (2) require disclosure of all author roles and contributions, and (3) specify that gift or honorary authorship, coercive authorship, ghost authorship, and omitting authors who have met the articulated standards are always unacceptable. Societies and journals should work expeditiously to develop such standards in disciplines that do not already have them.**

Authorship practices are a fundamental component of the research enterprise's operation, and observance of good practices is a key factor in ensuring research integrity. Authorship crucially designates who bears responsibility for the work. By communicating the assumptions made and methods used in conducting experiments, researchers allow others to replicate, extend, and where necessary,

correct their work. As a result, science is typically a cumulative exercise that produces a growing body of reliable knowledge. Clarifying authorship responsibility is also critical in case of error or allegations of misconduct.

Detrimental practices such as coercive authorship, gift authorship, and unacknowledged ghost authorship impair the usefulness and reliability of authorship as the central institution for assigning credit for reported work, fixing responsibility for that work's quality and integrity, and communicating critical information that allows other researchers to replicate, extend, and where necessary, correct that work.

Although some disciplines have developed clear authorship guidelines, authorship practices and conventions are largely left to individual institutions and journals. Greater clarity at the disciplinary level about the significant intellectual contributions that merit authorship, the roles that do not merit authorship, the significance of author order, and the responsibilities of a primary or corresponding author would be very helpful in facilitating appropriate decisions and practices in labs and collaborations. Universal condemnation (i.e., by all disciplines) of gift or honorary authorship, coercive authorship, and ghost authorship would also contribute to changing the culture of research environments where these practices are still accepted. Universal adoption of the requirement that all authorship roles be disclosed, as is the case for a growing number of journals, and commitment to the principle that all contributors who merit authorship should be listed would also be positive steps.

The committee favors an approach that authorship should be established through a significant intellectual contribution to the work in at least one area, such as planning, performing, analyzing, or writing. All authors should have the opportunity to approve the final manuscript.

The committee recognizes that flexibility in the development and implementation of authorship guidelines is needed due to significant differences between disciplines.

**RECOMMENDATION SIX: Through their policies and through the development of supporting infrastructure, research sponsors and science, engineering, technology, and medical journal and book publishers should ensure that information sufficient for a person knowledgeable about the field and its techniques to reproduce reported results is made available at the time of publication or as soon as possible after that.**

The information needed to verify and build upon published results can vary by field and discipline. Examples of such information include the specification of agents, materials and reagents, digital data, and software code and scripts used for analysis and production of results. With new advances in technology, such as the wide availability of image manipulation software as well as the pervasive ap-

plication of statistical and computational methods, continued adaptation and development of reporting standards and scientific practices are essential. In almost every area of scientific research, researchers have been quick to adopt advances in technology to accelerate the progress of research activities. In some specific cases discussed in the report, the absence of standards and lack of adherence to best practices have enabled fabricated work to go undetected or uncorrected for long periods of time. As information technology advances continue to transform scientific methods, the development and wide implementation of best dissemination practices needs to keep pace.

A research process that uses computational tools and digital data introduces myriad new potential sources of error: Were the methods described in the paper transcribed correctly into computer code? What were the parameter settings, input data, and function invocation sequences? How were the raw data filtered and prepared for analysis? Can the figures and tables reported in the published article be replicated by the associated data and code? Access to the data and code that produced the results is paramount, both for replication and validation purposes and for reconciling any differences in independent implementations.

Computation has facilitated vastly greater complexity in research. For example, the number of computational steps in deriving a scientific finding can be enormous, and these steps may not be completely captured in the traditional methods section of a scientific publication. All details of the specific computations that generated results, encapsulated in the code and data, must be made available to others for the findings to be reproducible.

The experimental and computational protocols and detailed methodology relevant to reproducibility should be made available by researchers. These include digital objects such as raw data—in fields where raw data are digital—and software, including source codes, scripts, and code books, sufficient to enable replication of computational research findings by one skilled in the discipline. These should be made openly available and reusable at the time of publication and persistently linked to or embedded in research articles.

Responding to recent attention to the problem of reproducibility, the research enterprise is beginning to take important steps. Some journals have begun to implement requirements that authors make the data and computer code required to regenerate the published results available upon request (*Science*, 2011). Many universities and funding agencies have created online repositories to support the dissemination of digital data, and best practices promulgating the routine sharing of digital scholarly objects that support verification of published findings must continue across computational research. Current digital data practices vary significantly by field and discipline, and making certain types of data broadly accessible presents special challenges. For example, the need to ensure privacy and anonymization of personal and clinical data that are to be shared requires technical ingenuity and imposes real costs. The successful development and implementation of new standards and requirements will depend upon sufficient

investments in necessary human and physical infrastructure, as described below in Recommendation Seven. Some fields and disciplines, such as astronomy, provide positive examples in which large amounts of digital data are being made widely available, while the infrastructure needs of other fields remain significant (NAS-NAE-IOM, 2009a).

The Transparency and Openness Promotion Guidelines developed by the Center for Open Science constitute an important contribution that can be studied and adapted by various fields. As is the case with other tasks identified in this report, building a more accountable and transparent research enterprise is a long-term, multistakeholder challenge.

There are notable exceptions to the presumption that all data and code should be shared, such as human subject privacy protections. When these safeguards are not at issue, the scientific community has an opportunity to act on the Office of Science and Technology Policy's memorandum of February 2013 aimed at expanding access to federally funded research (Holdren, 2013).

Previous reports have made similar recommendations about access to data and code in broad and specific contexts (Fienberg et al., 1985; IOM, 2015; NAS-NAE-IOM, 2009a; NRC, 2003). While many of the steps necessary for implementing the recommendations contained in these previous reports remain to be taken, there is reason to hope that the importance of access to data and code is becoming sufficiently well recognized to enable significant progress in the next few years.

Massive national investments are being made in digital data collection, which is typically not hypothesis driven but is undertaken because it is possible. This opens potential new research questions, but it also heightens the imperative for data availability to enable the production of reliable scientific findings.

Finally, with automatic plagiarism detection software increasingly being used on published articles and research, it has become apparent that some misconduct can be caught prior to publication. The committee encourages publishers to coordinate knowledge and efforts to adopt new technologies as they become available to detect and reduce plagiarism prior to publication.

**RECOMMENDATION SEVEN: Federal funding agencies and other research sponsors should allocate sufficient funds to enable the long-term storage, archiving, and access of datasets and code necessary for the replication of published findings.**

Preparing data and code for release can be expensive and time-consuming. Researchers are currently rewarded for manuscript publication, but the professional rewards for preparing data and code for publication are minimal. The resources to support the endeavor are also often limited, and the feasibility and time required depend very much on the type of research data and how they were collected. This has the effect of penalizing those who spend the necessary time and resources to prepare data and code for publication. One way to address this

problem is for the community to adopt new practices, and recent changes in federal policy provide such an opportunity (Holdren, 2013).

In addition, journals should update their publication requirements to include access to data and codes needed to replicate results. These data and codes can be deposited at any repository that can reasonably guarantee a persistent URL, which should be provided in the text of the published paper. Even when complex computational architectures have been used that make independent execution of the software difficult, sharing the code openly allows others to inspect, assess, and perhaps adapt the methods.

The barriers to data sharing at the scale recommended here are significant, and this recommendation will take some time to implement. Setting priorities and achieving the necessary funding levels will require time. Efforts to make data available and encourage reproducibility should catalyze the development of new data tools that ultimately reduce costs over time. The key is to ensure that data are in a long-term repository with metadata that allow third parties to reuse them.

To facilitate the reuse of scientific code and data, these objects should be shared in such a way as to maximize access while respecting scientific norms such as attribution (Stodden, 2009). Permissive open licensing, such as the MIT License or Modified BSD License for software or the Creative Commons Public Domain certification for data, should be used.

**RECOMMENDATION EIGHT: To avoid unproductive duplication of research and to permit effective judgments on the statistical significance of findings, researchers should routinely disclose all statistical tests carried out, including negative findings. Research sponsors, research institutions, and journals should support and encourage this level of transparency.**

Available evidence indicates that scientific publications are biased against presenting negative results and that the publication of negative results is on the decline (Fanelli, 2010, 2012). In extreme cases, where nearly identical experiments are run a number of times with one positive result being reported and multiple negative results discarded, the failure to report negative results constitutes a detrimental research practice (Couzin-Frankel, 2013). Yet, in recent years, several analyses and opinion pieces have pointed to the value of publishing negative results. For example, dissemination of negative results has prompted a questioning of established paradigms, leading ultimately to groundbreaking new discoveries (Anderson et al., 2013). Publication of negative results can also lead to the uncovering of flaws and the subsequent development of improved research methods. For example, a number of papers reported the negative results of work seeking to replicate research on vaccines and autism discussed in Chapter 5 and Appendix D.

Changing the culture of research and publication so that negative results

reporting is expected and replication efforts are valued will require a persistent effort on the part of disciplines, sponsors, and journals. The attention received by several recent replication efforts, one of which involved publication in *Science*, is encouraging (OSC, 2015). As routine reporting of negative results and statistical tests becomes the standard for all fields, research spending will become more productive and more knowledge will be generated per dollar of research investment.

***FINDING C: Improved strategies for fostering research integrity and for addressing threats to integrity posed by research misconduct and detrimental research practices need to be based on knowledge and evidence that does not currently exist. Investments are needed in research that improves understanding of key issues such as the relationship between structural conditions in science and the tendency for individuals to practice research according to the values and norms of integrity or to deviate from those values and norms. Improving knowledge in this area is essential to the long-term health of the research enterprise itself.***

Upholding the values of objectivity, honesty, and openness in the contemporary context requires that research institutions, the federal government, science, engineering, technology, and medical journal and book publishers, and scientific societies should, individually and in collaboration, examine these systemic conditions and their impacts on incentives. This research needs to bring to bear the best of what is known about influences on human decision making from a range of social science fields to guide actions to improve research climates so that they reflect and reinforce the core values of science. This research should complement the research environment assessment activities at institutions discussed above in Recommendation Two. These investments should be directed to helping all participants in the research enterprise, from the local to the international, act upon the findings to reinforce the values and norms underlying integrity. National and potentially international benchmarks are needed so that assessments can be understood in light of disciplinary differences, much as national medical cost benchmarking is shedding light on where some unnecessary procedures are being undertaken in some regions, with corrective actions being taken.

**RECOMMENDATION NINE: Government agencies and private foundations that support science, engineering, and medical research in the United States should fund research to quantify, and develop responses to, conditions in the research environment that may be linked to research misconduct and detrimental research practices. These research sponsors should use the data accumulated to monitor and modify existing policies and regulations.**

Material presented in Chapters 5 and 6 illustrates that understanding of the causes and incidence of research misconduct and detrimental research practices has increased but that critical knowledge gaps remain. For example, official statistics on findings of research misconduct may represent a lower bound on incidence, with survey data pointing to a significantly higher incidence of misconduct, but no reliable estimate of incidence or trends exists. Also, detrimental research practices are more widespread and may ultimately be more damaging to the research enterprise than research misconduct, which points to the need to address challenges to research integrity more broadly. In addition, trends in some indicators—such as declining success rates for grant applications, and an increasing ratio of PhD production to available faculty positions—raise the possibility that both local organizational environments and the broader structural arrangements of research are moving in directions that might threaten research integrity. Additional theoretically grounded research with subsequent testing in practice is warranted to more completely inform efforts to improve research environments and incentive structures.

Data generated through regular institutional research integrity assessments (discussed under Recommendation Two), research on the factors contributing to research misconduct and how to address them, and information on effective educational approaches could provide valuable input to the policies and practices of research sponsors and federal agencies charged with overseeing institutional research misconduct investigations. For example, the RCR education policies of the National Science Foundation and National Institutes of Health should be modified over time as knowledge improves. Also, a better understanding of the linkages between hypercompetitive research environments and misconduct and detrimental practices could help to support and inform changes. Where elements are identified that support particularly robust integrity environments, they should be broadly shared.

Evidence could inform better policy in a number of other areas. For example, although the Office of Research Integrity and the National Science Foundation's Office of the Inspector General both follow the 2000 federal research misconduct policy, there are a number of clear differences between the agencies in how they implement the policy, as discussed in Chapter 7. These include differences in how those found to have committed research misconduct are publicly identified, the scope of action available to the agencies outside formal investigations, and regulatory relationships between the agencies and research institutions. Greater understanding of the impacts of such differences could be helpful in determining whether and how to harmonize approaches across the federal government. Research bearing on other issues, such as what sort of corrective actions are appropriate for those who have committed research misconduct, how offenders should be rehabilitated, and the possible positive impacts on research integrity associated with data and code access mandates, would also be very useful to policy makers.

Finally, current practices in research funding and organization may contribute to a higher incidence of research misconduct and detrimental research practices, at least in some disciplines and institutions. Addressing the underlying structural problems of funding and organization would require significant policy changes that go beyond the scope of this study. However, a better understanding of the linkages between hypercompetitive research environments and misconduct and detrimental practices could help to support and inform such changes.

**RECOMMENDATION TEN: Researchers, research sponsors, and research institutions should continue to develop and assess more effective education and other programs that support the integrity of research. These improved programs should be widely adopted across disciplines and across national borders.**

Formal responsible conduct of research education and training efforts can play an important role in fostering integrity and strengthening research environments. Evidence developed to date indicates that much remains to be learned about the approaches that are most effective. RCR education should be looked to as a key element in strategies to promote integrity, but perhaps not as a primary means of addressing research misconduct and detrimental research practices in the short term. Evidence-based assessment and improvement of RCR education programs is needed, with the focus expanded to include the social and institutional environment for research. RCR education should engage not only junior scientists but also senior research scientists and industrial researchers.

***FINDING D: Working to ensure research integrity at the global level is essential to strengthening science both in the United States and internationally.***

The research enterprise is increasingly global in nature, and an international focus is imperative when seeking improvements in systems for safeguarding research integrity. As illustrated in recent media reports, all countries that perform a significant amount of research have experienced challenges in the area of research integrity, including high-profile cases of misconduct and, often, deficiencies in institutional and governmental responses. The World Conferences on Research Integrity have helped build a global community of experts. These conferences and other activities, such as Organisation for Economic Co-operation and Development workshops held to develop standard contract language for use in international collaborations, have made clear the value of cross-border exchange and learning. At the same time, varying policy contexts related to research support and institutional oversight may make thorough global harmonization of policies and practices difficult or impossible, at least in the immediate future.

**RECOMMENDATION ELEVEN: Researchers, research institutions, and research sponsors that participate in and support international collaborations should leverage these partnerships to foster research integrity through mutual learning and sharing of best practices, including collaborative international research on research integrity.**

While the committee has put its primary focus on how to better foster research integrity in the United States, the study was informed by a changing global context. The problems of research misconduct and detrimental research practices, and their resulting negative impacts, are global. Lack of training or ineffective training of students abroad has an impact in the United States when students or faculty move to U.S. institutions or collaborate with U.S.-based researchers. Many of the most visible and publicized cases of research misconduct have involved international coauthorship. Several recent research misconduct investigations undertaken by research institutions outside the United States have been exemplary and illustrate that U.S. researchers, institutions, and sponsors can learn a great deal from international colleagues (Ishii et al., 2014; Levelt et al., 2012).

In addition, disciplinary differences in research practices often vary to a greater degree than do differences between countries. Disciplines are for the most part global in scope, and disciplinary efforts to examine and upgrade practices will tend to be global as well. The Levelt report of the investigation of Diederik Stapel's research misconduct by the three institutions that employed him over a period of several decades identified a number of weaknesses in practices that have been widely tolerated in social psychology, performing a long-term service for all researchers in this field.

Also, researchers and institutions have many opportunities to learn from each other. Just as U.S. institutions can learn from other U.S. institutions that are more effective at education or at addressing allegations of misconduct, mutual learning between U.S. and overseas institutions can encourage improvement and diffusion of best practices.

Given that research misconduct, detrimental research practices, and the need to foster research integrity are challenges facing all countries that fund and perform research, the global research enterprise will benefit from the knowledge gained from the research agenda outlined under Finding C, above. Expanding this research agenda to a global scale would be beneficial to all. For example, exploration of cross-national or cross-cultural differences in attitudes and norms related to research behaviors (e.g., plagiarism) would be useful input to the development of targeted educational interventions. Development of a global evidence base on research integrity could accelerate the diffusion of effective approaches to addressing specific problems or issues. The global interacademy organizations are playing a role in this process with their publications *Responsible Conduct in the Global Research Enterprise: A Policy Report* (IAC-IAP, 2012) and *Doing Global Science: A Guide to Responsible Conduct in the Global Research Enterprise* (IAP, 2016).

# Appendixes

# Appendix A

# Biographical Information on the Committee and Staff

## CHAIR

**Robert M. Nerem, Institute Professor and Parker H. Petit Professor Emeritus, Institute for Bioengineering and Bioscience, Georgia Institute of Technology**
Dr. Nerem joined Georgia Tech in 1987 as the Parker H. Petit Distinguished Chair for Engineering in Medicine. He is an Institute Professor Emeritus, and he was the founding Director of the Parker H. Petit Institute for Bioengineering and Bioscience, a research institute established in 1995 to bring biochemistry, bioengineering, and biology faculty together so as to create a "convergent," interdisciplinary culture. He also was the Director of the Georgia Tech/Emory Center for the Engineering of Living Tissues, a National Science Foundation-funded Engineering Research Center, from 1998 to 2009. Dr. Nerem received his Ph.D. in 1964 from Ohio State University and is the author of more than 200 publications. Over the years he has served the community in a variety of ways. This includes his extensive involvement with the International Federation for Medical and Biological Engineering (IFMBE), serving as President from 1988 to 1991 and being the Founding President of the International Academy for Medical and Biological Engineering from 1997 to 2000. He also was the President of the International Union for Physical and Engineering Sciences in Medicine from 1991 to 1994. Dr. Nerem was the Founding President and is a Fellow of the American Institute for Medical and Biological Engineering (AIMBE). He has served on the advisory boards of a number of companies including startups, and from 2000 to 2003 he was a member of the Food and Drug Administration Science Board. From 2003 to 2006 he was a part-time Senior Advisor for Bioengineering in the National Institute for Biomedical Imaging and Bioengineering at the National Institutes of Health. In 1988 Dr. Nerem was elected to the National Academy of Engineering

(NAE) and in 1992 to the Institute of Medicine of the National Academy of Sciences, what is now named the National Academy of Medicine. In 1994 he was elected a Foreign Member of the Polish Academy of Sciences, in 1998 a Fellow of the American Academy of Arts and Sciences, and in 2006 a Foreign Member of the Swedish Royal Academy of Engineering Sciences. In 2008 Dr. Nerem was selected by NAE for the Founders Award, and in 2011 he was made an IFMBE Honorary Life Member. In 2015 IFMBE selected Dr. Nerem for the inaugural John A. Hopps Award.

## MEMBERS

**Ann M. Arvin, Lucile Packard Professor of Pediatrics and Microbiology and Immunology, Stanford University School of Medicine; and Vice Provost and Dean of Research, Stanford University**

Ann Arvin, M.D., is the Lucile Salter Packard Professor of Pediatrics and Professor of Microbiology and Immunology, Stanford University School of Medicine, and the Vice Provost and Dean of Research, Stanford University. As Vice Provost, she oversees Stanford's 18 interdisciplinary institutes as well as university research policies, compliance with regulations concerning the responsible conduct of research and human and animal research, and the Office of Technology Licensing. Her laboratory research focuses on molecular mechanisms of varicella zoster virus infection and immune responses to this common human herpesvirus. Her clinical research seeks to improve the understanding of the developing immune system in infants and young children in the context of viral infections and vaccines. Her work has been recognized by election to the American Academy of Arts & Sciences, the National Academy of Medicine, the American Association for the Advancement of Science, the Association of American Physicians, and the American Pediatric Society. Her past and current national committee service includes the National Academy of Sciences/National Research Countil (NAS/NRC) Board on Life Sciences, the Director's Advisory Council of the National Institute of Allergy and Infectious Diseases, and NAS/NRC Committees including the Committee on Federal Research Regulations and Reporting Requirements, the Committee on Policy and Global Affairs, and the Committee on Science, Technology and Law. Dr. Arvin was chief of the Infectious Diseases Division of the Lucile Packard Children's Hospital at Stanford from 1984 to 2006. She received her A.B. from Brown University, M.A. in philosophy from Brandeis University, and M.D. degree from the University of Pennsylvania. She completed her residency in pediatrics at the University of California, San Francisco, and subspecialty training in infectious diseases at UCSF and Stanford University.

**Rebecca M. Bergman, President, Gustavus Adolphus College**

Ms. Rebecca M. Bergman serves as the President of Gustavus Adolphus College, a liberal arts college located in St. Peter, MN. Ms. Bergman served as Vice

President, Research and Technology for Cardiac Rhythm Disease Management (CRDM) of Medtronic, Inc. from 2012 to 2014. Ms. Bergman served as a Vice President of New Therapies & Diagnostics, Cardiac Rhythm Disease Management at Medtronic, Inc., from January 2009 until 2012. Ms. Bergman served as a Vice President for Science & Technology of Medtronic, Inc., from 2002 until January 2009. Ms. Bergman has more than 26 years of experience in the medical technology industry including over 18 years of experience in research and technology management and product development. She has been an Independent Director of Sigma-Aldrich Corporation since May 2008. She serves as a Director of Sigma-Aldrich Company Limited. She served as a Director of TEI Biosciences Inc. Ms. Bergman has served as a member of the National Advisory Council of the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health and serves on a number of academic advisory boards. Ms. Bergman is a Fellow of the American Institute for Medical and Biological Engineering and a member of the National Academy of Engineering. She has a B.S. in chemical engineering from Princeton University and an honorary Ph.D. in engineering from Drexel University. She has also completed graduate studies in chemical engineering and material science at the University of Minnesota.

**Moses Chan, Evan Pugh Professor of Physics, Pennsylvania State University**
Moses Chan is Evan Pugh Professor of Physics at Pennsylvania State University. He is an alumnus of Bridgewater College and Cornell University, where he earned his Ph.D. in 1974 and was a Postdoctoral Associate at Duke University. He has been a professor at Penn State's University Park Campus since 1979. Through the years, Professor Chan's work has spanned many diverse topics. For his numerous contributions to low-temperature physics, in 1996 he shared the prestigious Fritz London Memorial prize with Carl Wieman and Eric A. Cornell. He was elected a member of the National Academy of Sciences in 2000, and a Fellow of the American Academy of Arts & Sciences in 2004.

**C. K. Gunsalus, Director, National Center for Professional and Research Ethics, University of Illinois**
C. K. Gunsalus is the Director of the National Center for Professional and Research Ethics, Professor Emerita of Business, and Research Professor at the Coordinated Science Laboratory. She has been on the faculty of the Colleges of Business, Law, and Medicine at the University of Illinois at Urbana-Champaign and served a range of administrative roles in the campus administration as well as Special Counsel in the Office of University Counsel. She was the Principal Investigator on the National Science Foundation–funded cooperative agreement that provided $1.5M to initiate Ethics CORE, the National Online Ethics Resource Center, and a recent $2.7M award focusing on the leadership needs of the research university of the future. Her work focuses on topics such as leadership development, ethics, research integrity, whistleblowing, and professionalism.

She was a member of the U.S. Commission on Research Integrity and served for 4 years as chair of the American Association for the Advancement of Science (AAAS) Committee on Scientific Freedom and Responsibility. She is an elected Fellow of the AAAS. She is the author of two books published by the Harvard University Press and a range of publications on institutional and research integrity.

**Deborah G. Johnson, Anne Shirley Carter Olsson Professor of Applied Ethics Emeritus, University of Virginia**

Deborah G. Johnson is the Anne Shirley Carter Olsson Professor of Applied Ethics Emeritus in the Science, Technology, and Society Program in the School of Engineering and Applied Sciences at the University of Virginia. Best known for her work on computer ethics and engineering ethics, Johnson's research examines the ethical, social, and policy implications of technology and engineering, especially information technology. Johnson is the author/editor of nine books including four editions of *Computer Ethics* (1985, 1991, 2001, 2009) and her most recent book, *Technology & Society: Engineering Our Sociotechnical Future* with J. M. Wetmore (2009). She has published over 100 papers in a variety of journals and edited volumes. Her research has repeatedly received support from the National Science Foundation. Most recently she received funding for a project on *Surveillance and Transparency as Sociotechnical Systems of Accountability* (2010-2012) and another project on *Ethics for Developing Technologies: An Analysis of Artificial Agents* (2011-2013). During 1992-1993 she was a Visiting Professor in the Department of Civil Engineering and Operations Research of Princeton University where she worked on a National Science Foundation project on ethics and computer decision models. In 1994 and 1995 and again in 2000, she received National Science Foundation funding to conduct workshops to prepare undergraduate faculty to teach courses and course modules on ethical and professional issues in computing. Active in professional organizations, Johnson has served as President of the Society for Philosophy and Technology, President of the International Society for Ethics and Information Technology, Treasurer of the ACM Special Interest Group on Computers and Society, Chair of the American Philosophical Association Committee on Computers and Philosophy, and a member of the Executive Board of the Association for Practical and Professional Ethics. She has also served on several committees of the National Academies.

**Michael A. Keller, Ida M. Green University Librarian and Director, Academic Information Resources, Stanford University**

At Stanford, Michael A. Keller is the Ida M. Green University Librarian, Director of Academic Information Resources, Publisher of HighWire Press, and Publisher of the Stanford University Press. These titles touch on his major professional preoccupations: commitment to support of research, teaching, and learning; effective deployment of information technology hand-in-hand with materials; and

active involvement in the evolution and growth of scholarly communication. He may be best known at present for his distinctively entrepreneurial style of librarianship. As University Librarian, he endeavors to champion deep collecting of traditional library materials (especially of manuscript and archival materials) concurrent with full engagement in emerging information technologies. Uniquely, Keller's responsibilities at Stanford encompass libraries, cybraries, academic and residential computing, and publishing and publishing services. As a result of his work in collection development at Cornell, Berkeley, Yale, and Stanford—which provided broad exposure to the global publication and bookselling trades—Keller became convinced of a need for correction in the marketplace of scholarly communications, especially journal publishing. Long involved in the great debate on serials pricing, especially in the arenas of science, technology, and medicine, he has served as advisor, consultant, and committee member to the American Association for the Advancement of Science and other scholarly societies. Thus in 1995, in response to scholars' requests for assistance to their scholarly societies, he established the HighWire Press as an enterprise within the Stanford University Libraries to provide online co-publishing services initially to three scholarly journals. As of January 2009, HighWire Press has grown to support over 1,200 high-impact STM journals among more than 130 major scholarly societies, over 1.8 million articles of which are available free online. Keller was educated at Hamilton College (B.A. biology, music 1967), SUNY Buffalo (M.A., musicology, 1970), and SUNY Geneseo (M.L.S., 1971). From 1973 to 1981, he served as Music Librarian and Senior Lecturer in Musicology at Cornell University and then in a similar capacity at UC Berkeley. While at Berkeley, he also taught musicology at Stanford University and began the complete revision of the definitive Music Research and Reference Materials, an annotated bibliography popularly known as Duckles in honor of its original compiler. Yale called him to the post of Associate University Librarian and Director of Collection Development in 1986. In 1993, he joined the Stanford staff as the Ida M. Green Director of Libraries. In 1994, he was named to his current position of University Librarian and Director of Academic Information Resources. In 1995, by establishing HighWire Press, he became its publisher, and in April 2000, he was assigned similar strategic duty for the Stanford University Press. In 2010, Keller became an elected Fellow of the American Academy of Arts & Sciences.

**W. Carl Lineberger, E. U. Condon Distinguished Professor of Chemistry and Fellow of JILA, University of Colorado**

Carl Lineberger serves as E. U. Condon Distinguished Professor of Chemistry and Biochemistry at the University of Colorado Boulder. Dr. Lineberger was nominated by President Obama and confirmed by the U.S. Senate to be a Member of the National Science Board for the 2011-2016 term. In 2016, President Obama appointed him to serve a second 6-year term on NSB. Dr. Lineberger has chaired the National Science Foundation Advisory Committees on Mathematical and

Physical Sciences and on the Science and Technology Centers. He has completed service on the National Academy of Sciences/National Research Council Committee on Science, Engineering, and Public Policy; the NRC Governing Board; and the Department of Energy Committee on New Science for a Secure and Sustainable Energy Future. His work is primarily experimental, using a wide variety of laser-based techniques to study structure and reactivity of gas-phase ions. He has published 2,850 papers in major scientific journals, and his graduate students and postdoctoral associates hold major research-related positions throughout the world. He received his B.S., M.S., and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology.

**Brian C. Martinson, Senior Research Investigator, HealthPartners Institute, Core Investigator, Minneapolis VA, Center for Chronic Disease Outcomes Research, and Associate Professor, University of Minnesota, Department of Medicine**

Brian C. Martinson is a senior research investigator at HealthPartners Institute, Core Investigator, Minneapolis VA, Center for Chronic Disease Outcomes Research, and Associate Professor, University of Minnesota, Department of Medicine. He earned his Ph.D. in sociology and demography at the University of Wisconsin–Madison, and his postdoctoral training was in cardiovascular behavioral health at the University of Minnesota, Minneapolis. Over the past 15 years, Dr. Martinson has contributed both substantively and methodologically to improving the understanding of research-related behavior (both that which contributes to research integrity and that which can undermine it), as well as to understanding the determinants of such behavior. He has led or co-led four federally funded research projects on these topics. His work was the first to document surprisingly high levels of self-reported, undesirable research-related behavior in large samples of NIH-funded researchers (see, e.g. Martinson et al., 2005, *Nature*, "Scientists behaving badly"). This line of research was also among the first to document the potential influences of perceptions of organizational justice on research-related behaviors among academic researchers. As Co-Principal Investigator with Dr. Carol Thrush, he co-led the development of and assessment of the validity and reliability of a survey instrument to assess the integrity of organizational climates in research organizations, resulting in a tool called the Survey of Organizational Research Climate (SOURCE). Most recently, he served as PI of a two year project in the VA conducting a randomized controlled trial using the SOURCE tool as part of a project testing the efficacy of a reporting and feedback intervention to improve research integrity climates in VA research settings. In 2009–2010, Dr. Martinson served as a consultant to a three-university consortium participating in the U.S. Council of Graduate Schools' Project on Scholarly Integrity. During that same time frame, he served on an invited expert panel on research integrity, convened by the Council of Canadian Academies at the request of Industry Canada, leading to the report *Honesty, Accountability*

*and Trust: Fostering Research Integrity in Canada*. In 2014, he served on the planning committee and subsequently as a speaker at a workshop of the Roundtable on Science and Welfare in Laboratory Animal Use (An ILAR Roundtable Series)—"The Missing "R": Reproducibility in a Changing Research Landscape, National Academy of Sciences, Institute for Laboratory Animal Research. Washington, DC, June 4-5. In 2015, he was an invited participant (one of 20) in a 2-day colloquium, sponsored by the American Academy of Microbiology, focused on issues of reproducibility of research in that field, entitled "Promoting Ethical Practices in the Scientific Enterprise," Washington, DC, October 14-15.

**Victoria Stodden, Associate Professor of Information Sciences, University of Illinois at Urbana-Champaign**

Victoria Stodden joined the School of Information Sciences as an associate professor in Fall 2014. She is a leading figure in the area of reproducibility in computational science, exploring how we can better ensure the reliability and usefulness of scientific results in the face of increasingly sophisticated computational approaches to research. Her work addresses a wide range of topics, including standards of openness for data and code sharing, legal and policy barriers to disseminating reproducible research, robustness in replicated findings, cyberinfrastructure to enable reproducibility, and scientific publishing practices. Dr. Stodden co-chairs the National Science Foundation Advisory Committee for CyberInfrastructure and is a member of the NSF Directorate for Computer and Information Science and Engineering Advisory Committee. She also serves on the National Academies Committee on Responsible Science: Ensuring the Integrity of the Research Process; and the Data Science Post-Secondary Education Roundtable. Previously an assistant professor of statistics at Columbia University, Dr. Stodden taught courses in data science, reproducible research, and statistical theory and was affiliated with the Institute for Data Sciences and Engineering. She co-edited two books released in 2014—*Privacy, Big Data, and the Public Good: Frameworks for Engagement*, published by Cambridge University Press, and *Implementing Reproducible Research*, published by Taylor & Francis. Dr. Stodden earned her Ph.D. in statistics and her law degree from Stanford University. She also holds a master's degree in economics from the University of British Columbia and a bachelor's degree in economics from the University of Ottawa.

**Sara Wilson, Associate Professor, Mechanical Engineering, University of Kansas**

Sara Wilson joined the Department of Mechanical Engineering at the University of Kansas in 2001. In addition to her position as an associate professor in mechanical engineering, she is the academic director of the Bioengineering Graduate Program at the University of Kansas and has a courtesy appointment in physical therapy and rehabilitation sciences at the University of Kansas Medical Center. Prior to joining University of Kansas, she was a postdoctoral researcher at the

University of Virginia. Dr. Wilson conducts research in the neuromuscular control of human motion using engineering principles from control theory and dynamics. She has studied the effects of occupational exposures such as vibration on the lumbar spine and low back disorders. She is also involved in the development of medical devices used in physical therapy, obstetrics, and internal medicine. She is deputy editor of the *Journal of Applied Biomechanics*. She was the 2015–2016 chair of the American Society of Mechanical Engineers Bioengineering Division. She is also active in teaching and development of educational tools in the area of responsible conduct of research for graduate students in engineering. She was a 2006 W. T. Kemper Fellow for Teaching Excellence at the University of Kansas. Dr. Wilson received her Ph.D. in medical engineering from Massachusetts Institute of Technology in 1999, her master's degree in mechanical engineering from Massachusetts Institute of Technology in 1994, and a bachelor's degree in biomedical engineering from Rensselaer Polytechnic Institute in 1992.

**Paul Root Wolpe, Asa Griggs Candler Professor of Bioethics; Director, Center for Ethics, Emory University**

Paul Root Wolpe, Ph.D., is the Asa Griggs Candler Professor of Bioethics, the Raymond F. Schinazi Distinguished Research Chair in Jewish Bioethics, a Professor in the Departments of Medicine, Pediatrics, Psychiatry, and Sociology, and the Director of the Center for Ethics at Emory University. Dr. Wolpe also serves as the first Senior Bioethicist for the National Aeronautics and Space Administration (NASA), where he is responsible for formulating policy on bioethical issues and safeguarding research subjects. He is Co-editor of the *American Journal of Bioethics* (*AJOB*), the premier scholarly journal in bioethics, and Editor of *AJOB Neuroscience*, and sits on the editorial boards of over a dozen professional journals in medicine and ethics. Dr Wolpe is a Past President of the American Society for Bioethics and Humanities; a Fellow of the College of Physicians of Philadelphia, the country's oldest medical society; a Fellow of the Hastings Center, the oldest bioethics institute in America; and was the first National Bioethics Advisor to Planned Parenthood Federation of America. Dr. Wolpe moved to Emory University in the summer of 2008 from the University of Pennsylvania, where he was on the faculty for over 20 years in the Departments of Psychiatry, Sociology, and Medical Ethics. He was a Senior Fellow of Penn's Center for Bioethics, and directed the Scattergood Program for the Applied Ethics of Behavioral Health and the Program in Psychiatry and Ethics at the School of Medicine. Dr. Wolpe is the author of over 125 articles, editorials, and book chapters in sociology, medicine, and bioethics, and has contributed to a variety of encyclopedias on bioethical issues. A futurist interested in social dynamics, Dr. Wolpe's work focuses on the social, religious, ethical, and ideological impact of technology on the human condition. Considered one of the founders of the field of neuroethics, which examines the ethical implications of neuroscience, he also writes about other emerging technologies, such as genetic engineering, nanotechnology, prosthetics,

and new reproductive technologies. His teaching and publications range across multiple fields of bioethics and sociology, including death and dying, genetics and eugenics, sexuality and gender, mental health and illness, alternative medicine, and bioethics in extreme environments such as space.

**Levi Wood, Assistant Professor, Georgia Institute of Technology**
Dr. Wood joined Georgia Tech as an assistant professor in August 2015. Prior to his current appointment, he was a postdoctoral fellow at the Beth Israel Deaconess Medical Center, Massachusetts General Hospital, and Harvard Medical School. There he used systems biology to elucidate novel signaling mechanisms in Alzheimer's disease and intestinal inflammation. Dr. Wood received his Ph.D. in mechanical engineering at the Massachusetts Institute of Technology, where he developed and used a microfluidic platform to identify dominant mechanisms governing vascular geometry during early vascular growth.

## STAFF

**Tom Arrison** is a program director in the Policy and Global Affairs division of the National Academies of Sciences, Engineering, and Medicine. He joined the Academies in 1990 and has directed a range of studies and other projects in areas such as international science and technology relations, innovation, information technology, higher education, and strengthening the U.S. research enterprise. Arrison is also the executive director of the InterAcademy Council/InterAcademy Partnership for Research. IAC produces reports on scientific, technological, and health issues related to the great global challenges of our time, providing knowledge and advice to national governments and international organizations. He earned M.A. degrees in public policy and Asian studies from the University of Michigan.

**Nina Ward** is a research associate in the Policy and Global Affairs (PGA) division at the National Academies of Sciences, Engineering, and Medicine. Ward supports PGA research efforts for Development, Security, and Cooperation (DSC) and the InterAcademy Partnership for Research. She has also formerly supported the Board on Higher Education and Workforce and the Committee on Women in Science, Engineering, and Medicine. She earned a B.A. in anthropology at Elon University and is currently pursuing an M.P.P. at the University of Maryland School of Public Policy.

**Dr. Lida Anestidou** is senior program officer at the Institute for Laboratory Animal Research of the U.S. National Academy of Sciences, where she directs a diverse portfolio of studies on the use of laboratory animals; biodefense and biosecurity; and research integrity/responsible conduct of research. Prior to this position she was faculty at the Center for Biomedical Ethics and Society, Vanderbilt

University Medical Center. She earned her doctorate in biomedical sciences from the University of Texas at Houston. Working with physiologist Norman Weisbrodt, she explored the effects of nitric oxide on the motility of the gastrointestinal musculature. Working with research integrity expert and biomedical ethics educator Elizabeth Heitman, she concurrently pursued her interests in biomedical ethics, scientific integrity, and science policy. Dr. Anestidou also holds a Doctor of Veterinary Medicine degree from Greece (her home country) and an M.S. in Veterinary Sciences from the University of Florida. She is an editorial board member of *Science and Engineering Ethics*, *Lab Animal*, and *SciTech Lawyer* and an ad hoc reviewer for the *American Journal of Bioethics*. She is a member of the National Conference of Lawyers and Scientists. Dr. Anestidou serves as an expert reviewer in the Ethics Evaluation of grant applications to the 7th Framework Program of the European Research Council and the European Commission Directorate General Research.

**Neeraj Prasad Gorkhaly** is an associate program officer at the National Academies of Science, Engineering and Medicine. Currently he works on the Board of Physics and Astronomy, and the National Materials and Manufacturing Board. Previously, he served in various capacities for the National Academies' Committee on Science, Engineering and Public Policy as well as the Board on Global Science and Technology. In the past decade he has participated in over 60 studies, reports, and workshops providing advice to the U.S. government on various scientific issues and policies, including the Norman Augustine–chaired report *Rising Above the Gathering Storm: Energizing and Employing America for a Brighter Economic Future*. He is also the president and founder of the Gorkhaly Foundation, a volunteer nonprofit organization based in Virginia, implementing social and economically sustainable projects in rural areas of Nepal. He is a graduate of the Ohio State University and a past fellow of the John Glenn Institute for Public Service and Public Policy.

**Maria Lund Dahlberg** is an associate program officer with the National Academies of Sciences, Engineering, and Medicine. She works with a number of groups across the institution, including the Board on Higher Education and Workforce, the central Office of Communications, and the National Academy of Medicine. She came to the National Academies by way of a Christine Mirzayan Science and Technology Policy Fellowship, which she received after completing all requirements short of finalizing the dissertation for her doctorate in physics at the Pennsylvania State University. Ms. Dahlberg holds a B.A. in physics from Vassar College and an M.S. in physics from the Pennsylvania State University.

**Steve Olson** has been a consultant writer since 1979 for the National Academy of Sciences, the National Academy of Engineering, the Institute of Medicine, the National Research Council, the President's Council of Advisors on Science and

Technology, the Office of Science and Technology Policy, the Howard Hughes Medical Institute, the National Institutes of Health, the National Science Foundation, and other organizations. He is the author of *Mapping Human History: Genes, Race, and Our Common Origins*, which was one of five finalists for the 2002 nonfiction National Book Award; *Count Down: Six Kids Vie for Glory at the World's Toughest Math Competition*; and *Eruption: The Untold Story of Mount St. Helens*, which was shortlisted for the Boardman Tasker Prize for Mountain Literature. He also has written for the *Atlantic Monthly*, *Science*, the *Smithsonian*, *Scientific American*, *Wired*, the *Yale Alumni Magazine*, the *Washingtonian*, *Slate*, *Astronomy*, *Science 82-86*, and many other magazines. From 1989 through 1992 he served as Special Assistant for Communications in the White House Office of Science and Technology Policy. He earned a bachelor's degree in physics from Yale University in 1978.

# Appendix B

# Agendas of Committee Meeting Public Sessions[1]

**First Meeting: March 18–20, 2012**
**Washington, DC 20001**

**AGENDA**

Monday, March 19, 2012

**1:00 PM** **Discussion of Study Goals with Sponsors**

Joel Kupersmith, Chief Research and Development Officer, Office of Research and Development, Department of Veterans Affairs
Patrick Glynn, Senior Technical Policy Advisor, Office of the Deputy Director for Science Programs, Office of Science, Department of Energy
Linda Gundersen, Director, Office of Science Quality and Integrity, U.S. Geological Survey, Department of the Interior

[1] In addition to the open sessions in which the committee heard from outside experts, the study process also included closed sessions during the meetings listed here, plus several meetings and numerous conference calls toward the latter part of the process that only involved committee members and staff.

James Kroll, Head of Administrative Investigations, Office of the Inspector General, National Science Foundation
John Galland, Director, Division of Education and Integrity, Office of Research Integrity, Department of Health and Human Services

**3:15 PM** **Discussion with Invited Experts**

Carrie Wolinetz, Associate Vice President for Federal Relations, Association of American Universities
Heather Pierce, Senior Director, Science Policy and Regulatory Counsel, Association of American Medical Colleges
Francesca Grifo, Senior Scientist and Director, Scientific Integrity Program, Union of Concerned Scientists
Ivan Oransky, Executive Editor, Reuters Health, and Co-Founder, Retraction Watch

**5:00 PM** **Comments from Other Experts and the Public (if needed)**

**Second Meeting: July 8–10, 2012**
**Palo Alto, California**

**AGENDA**

Monday, July 9, 2012

**8:45 AM** **Challenges and Tasks for Scientific Journals in Ensuring Research Integrity**

Drummond Rennie, UCSF (deputy editor, *JAMA*)
Donald Kennedy, Stanford University (former editor, *Science*)
Philip Campbell, *Nature* (by videoconference)

**10:45 AM** **Learning from the Duke Case and the IOM Translational Omics Report**

Keith Baggerly, MD, Anderson Cancer Center
Gilbert Omenn, University of Michigan
Robert Califf, Duke University Medical Center (by videoconference)

**12:30 PM** **Working Lunch: IOM Conflict of Interest Study**

Bernard Lo, Greenwall Foundation

**2:00 PM** **Digitization and New Scientific Methods: Implications for Research Integrity**

Mark Liberman, Penn
David Donoho, Stanford University
Sergey Fomel, University of Texas at Austin

**4:00 PM** **Lessons and Experiences from the Project *A Collegial Defense Against Irresponsible Science***

Joan Sieber, California State University–East Bay

**5:00 PM** **The Federal Research Misconduct Definition**

Arthur Bienenstock, Stanford University

**6:30 PM** **Dinner: Perspectives on Research Integrity**

Keith Yamamoto, UCSF

**Third Meeting: August 14–15, 2012**
**Washington, DC**

**AGENDA**

Tuesday, August 14, 2012

**8:40 AM** **Industry Perspectives on Research Integrity—Part One**

Ellen Williams, BP (by videoconference)
Gillian Woollett, Avalere Health LLC

**9:45 AM** **Research Misconduct: Discovery, Reporting, and Assessing Impacts**

Carolyn Phinney, Consultant, Counselor of Whistle-Blowers
Thomas Evans, Montana State University
Mary Allen, University of Colorado

**11:25 AM** **Institutional Perspective**

Claude Canizares, MIT (by videoconference)

**12:15 PM** **Working Lunch: Funder Perspectives**

Rod Ulane, NIH
James Kroll, NSF-IG (invited)
Susan Garfinkel, ORI
Brendan Godfrey, DOD

**2:30 PM** **Industry Perspectives on Research Integrity—Part Two**

Richard Kuntz, Medtronic
Mark Wegman, IBM

**3:35 PM** **Society and Association Effort to Foster Research Integrity**

Mark Frankel, AAAS
Daniel Denecke, CGS
Cathee Johnson Phillips, NPA

**Fostering Responsible Research: A Roundtable Discussion with Societies**

**December 13, 2012 (Thursday)**
**Washington, DC**

**AGENDA**

| | |
|---|---|
| **9:00 AM** | **Welcome, Introductions, and Update on the Responsible Science Study** |
| **9:20 AM** | **General Perspectives on Responsible Science** |
| **10:00 AM** | **Societies and Standards** |
| **11:00 AM** | **Society Participation in National Academies Studies** |
| **11:30 AM** | **Societies and Scholarly Communication** |
| **12:30 PM** | **Lunch** |
| **1:00 PM** | **Societies and Education, Training, and Mentorship** |
| **1:40 PM** | **Concluding Discussion and Possible Next Steps** |
| **2:30 PM** | **Adjourn** |

# Appendix C

# Assessing the Effectiveness of Responsible Conduct of Research Training: Key Findings and Viable Procedures[1]

*Michael D. Mumford*
*The University of Oklahoma*

**ABSTRACT**

*Of the many interventions that might be used to improve the responsible conduct of research, educational interventions are among the most frequently employed. However, educational interventions come in many forms and have proven of varying effectiveness. Recognition of this point has led to calls for the systematic evaluation of responsible conduct of research educational programs. In the present effort, the basic principles underlying evaluation of educational programs are discussed. Subsequently, the application of these principles in the evaluation of responsible conduct of research educational programs is described. It is concluded that systematic evaluation of educational programs not only allow for the appraisal of instructional effectiveness but also allows for progressive refinement of educational initiatives.*

Ethics in the sciences and engineering is of concern not only because of its impact on progress in the research enterprise but also because the work of

[1] As the committee launched this study, members realized that questions related to the effectiveness of Responsible Conduct of Research education programs and how they might be improved were an essential part of the study task. A significant amount of work has been done to explore these topics. This work has yielded important insights, but additional research is needed to strengthen the evidence base relevant to several key policy questions. The committee asked one of the leading researchers in this field, Michael D. Mumford, to prepare a review characterizing the current state of knowledge and describing future priorities and pathways for assessing and improving RCR education programs. The resulting review constitutes important source material for Chapter 10 of the report. The committee also believes that the review adds value to this report as a standalone document, and is including it as an appendix.

scientists and engineers impacts the lives of many people. Recognition of this point has led to a number of initiatives intended to improve the ethical conduct of investigators (National Academy of Engineering, 2009; Institute of Medicine and National Research Council, 2002; National Academy of Sciences, National Academy of Engineering, and Institute of Medicine, 1992). Although a number of interventions have been proposed as a basis for improving ethical conduct, for example, development of ethical guidelines, open data access, and better mentoring, perhaps the most widely applied approach has been ethics education (Council of Graduate Schools, 2012)—an intervention often referred to as training in the responsible conduct of research (RCR).

When one examines the available literature on RCR training, it is apparent that a wide variety of approaches have been employed. Some RCR courses are based on a self-paced, online, instructional framework (e.g., Braunschweiger and Goodman, 2007). Other RCR courses involve face-to-face instruction over longer periods of time using realistic exercises and cases (e.g., Kligyte et al., 2008). Some R,CR courses focus on specific ethical issues (DuBois and Duecker, 2009) while others are based on general theoretical models of ethical conduct (Bebeau and Thoma, 1994). Some programs focus on ethics within a particular discipline (e.g., Major-Kincade et al., 2001). Other programs, however, take a cross-field, or multidisciplinary, approach (e.g., Mumford et al., 2008). Some programs seek to encourage analysis of ethical problems (e.g., Gawthrop and Uhlemann, 1992) while others seek to ensure appropriate ethical behavior (e.g., Drake et al., 2005).

The variety of educational approaches, approaches differing in content, instructional techniques, breadth, and objectives, broaches a question—a question fundamental to the present effort. What RCR programs work and how well do they work? Answers to these questions are important not only because they allow us to develop RCR programs of real value in improving ethics, but they also provide a basis for the progressive improvement of instructional practices. Attempts to answer these questions and improve RCR instruction must ultimately be based on systematic program evaluation efforts. Accordingly, our intent in the present effort is to examine the evaluation of RCR educational programs to both determine what we know about the effectiveness of instruction and how we might go about improving RCR instruction.

## EVALUATION

### Principles

Evaluation is intended to demonstrate change in an outcome of interest (Gottman, 1995) as a result of an intervention, or a package of interventions (Shadish et al., 2002) with respect to a certain set of objects (Yammarino et al., 2005). This definition of program evaluation is noteworthy because it has a number of implications for the design of viable evaluation studies, including

studies intended to appraise the effectiveness of RCR instruction. We will begin by examining each key attribute of this definition of evaluation in the context of RCR instruction.

In RCR instruction the intervention is the educational program to which students have been exposed. Instructional interventions, however, are inherently complex, involving multiple facets—content, the instructor, exercises, the setting, student preparation, and duration (Goldstein, 1986) to mention a few. Evaluation of the instructional interventions is possible only when those facets of instruction, the intervention, have been held constant or reasonably constant. Thus in evaluation of RCR instruction it is critical that a standardized, consistently executed, set of instructional practices be employed. Given the complexity of training interventions, however, interventions are typically conceived of as a class, or certain type of, intervention—for example, in-class versus online instruction.

Educational interventions, like interventions in general, are expected to have certain effects. The effects of RCR instruction might be on ethical decision making (Mumford et al., 2006), perceptions of ethical climate (Anderson, 2010), or knowledge (Heitman and Bulger, 2006). What should be recognized here is that the nature of the intervention will influence the effects one expects to observe. As a result, the measures used to appraise the effects of one instructional program may not be identical to the measures used to appraise the effects of another instructional program. Although a variety of measures may be used to appraise the effects of RCR instruction, it is critical the measures employed evidence adequate reliability and validity (Messick, 1995). Reliability, consistency in scores, is critical for demonstrating change. Validity allows inferences, substantively justified inferences, to be drawn with respect to the nature of the changes observed.

Our foregoing observations bring us to the next critical issue of concern in evaluation studies—how is change to be demonstrated. Although statistical considerations are of concern in demonstrating change (Gottman, 1995), successful demonstration of change ultimately depends on the design used in evaluation studies (Shadish et al., 2002). Broadly speaking, change can be demonstrated in two ways. First, one can show that a group exposed to the intervention differs from a group not exposed to the change intervention. Second, one can show that objects, often people, differed after exposure to the change intervention—a pre-post design. Of course, pre-post designs with no intervention controls can be, and perhaps should be, employed (Cook and Campbell, 1979). However, in evaluation studies the other concerns arise in assessing change. One concern pertains to whether these changes are maintained over time. The other concern pertains to whether changes observed transfer to other tasks or performance settings (Goldstein, 1986).

The fourth, and final, aspect of this definition of evaluation pertains to the objects where change is to be observed. In studies of training education, we commonly assume the critical object of concern is the students taking the class. However, in RCR instruction a variety of other objects might also be of concern

(Steneck and Bulger, 2007). For example, one might be concerned with laboratory practices. Alternatively, one might be concerned with department or institutional climate. These observations are noteworthy because they point to the need to consider both the objects of concern in RCR training and potentially objects operating at different levels of analysis (Yammarino et al., 2005).

## Training

Consideration of the principles sketched out above is of concern in virtually any evaluation effort—including evaluation of RCR instruction. By the same token some unique concerns do arise in the evaluation of training programs such as RCR instruction. The three critical unique concerns pertain to uses of evaluation data, sample/design, and evaluation measures. In the following section we will consider each of these issues in the context of RCR training.

The principle use of evaluation data is determining whether the RCR instructional program did result in change on the measures being used to appraise program effects. Put more directly, program evaluation tells us whether the program worked. In this regard, however, it is important to bear in mind not only whether change was observed but also how large the observed changes were. As a result, effect size estimates are commonly used in evaluation of training programs. In this regard, however, it is important to recognize that stronger inferences of program effectiveness are permitted when effects are observed in other settings—in the laboratory as well as the classroom.

Although evaluation data are needed for indicating whether change, sizeable change, has resulted from instruction, evaluation data are commonly used to address three other critical issues. First, evaluation data may be used to improve instructional processes. For example, if knowledge improves but not ethical decision making as a result of RCR instruction, it is feasible to argue that changes in instruction are needed. Second, evaluation data provide a basis for day-to-day program management. For example, if one instructor consistently produces weak effects and/or weaker effects than other instructors, perhaps remedial interventions are needed to improve instructor performance. Third, evaluation data are used to identify best practices or model instructional programs—instructional programs that should provide a basis for progressive refinement of the instructional system (Cascio and Aguinis, 2004).

Concerns with samples and design pertain to the number of participants, and the nature of the measures and design, needed to provide viable estimates of effect size. Pre-post test designs, as individual differences designs, typically require samples of 100 or more individuals to produce stable estimates of effect size. Comparisons of trained individuals to untrained groups typically require stable estimates of group means and standard deviations—a cell size of 25 individuals per group. In studies where these conditions cannot be met, it is possible either

to employ a broader array of measures to strengthen inferences or, alternatively, to employ qualitative procedures to appraise program effects.

With regard to evaluation design, a general tendency to employ a pre-post design with untrained controls is preferred. In organizations, however, training effects may be inadvertently disseminated to participants. Inadvertent dissemination, and expectations induced by dissemination, may call for inclusion of additional controls. Moreover, people bring to any educational experience background, personal characteristics, and a work history. As a result, it is common in training evaluation to consider a wider variety of control measures than is dictated by evaluation designs per se such as student characteristics (e.g., interest in ethics), climate for transfer (e.g., mentor or work group support), student intentions (e.g., ethical goals), prior educational experiences (e.g., earlier ethics education), and field or discipline (Baldwin and Ford, 1988; Fleishman and Mumford, 1989; Mumford et al., 2007; Colquitt et al., 2000).

In training evaluation a critical concern has been the nature of the measures that should be used to appraise instructional effectiveness. Over the years, a number of taxonomies of potential evaluation measures for training, as a general class of interventions, have been proposed (Kirkpatrick, 1978; Aguinis and Kraiger, 2009). However, these classifications of training evaluation measures were not developed with respect to ethics training. Broadly speaking, however, seven distinct classes of measures have been developed that might be used to evaluate ethics training.

The first class of measures reflects *performance*. The performance measures used in evaluation of ethics instruction do not focus on real-world ethical performance or breaches in ethical conduct in part because of the frequency of such events and in part because of ethical concerns attached to measuring such events. Rather, to assess performance, low-fidelity simulation measures are used (Motowidlo et al., 1990). On low-fidelity simulations, people are presented with scenarios where an event has occurred that requires an ethical decision to be made. Multiple alternative responses to this scenario are presented where response options vary in ethicality. The available evidence indicates that well-developed ethical decision-making measures evidence adequate reliability and good construct validity (Mumford et al., 2006). For example, in the Mumford et al. (2006) study, poor decisions were found to be positively related to narcissism and negatively related to ethical conduct by major professors. With regard to these measures, however, coverage of relevant aspects of ethical decisions (e.g., decisions involving conflicts of interest or decisions involving authorship) must be considered. Moreover, as low-fidelity simulations, ethical decision-making measures are more appropriate when developed to be applicable to the field or discipline in which the person is working. Thus Mumford et al. (2006) developed ethical decision-making measures applying in the biological, health, and social studies, while Kligyte et. al., (2008) developed ethical decision-making measures for the engineering and physical sciences.

The second set of measures commonly used to appraise RCR instruction focuses on *knowledge*. Knowledge measures are typically intended to assess either recognition or recall of factual information presented in RCR training. Typically, a knowledge item presents a question where answers require recall of information provided in training. Valid and reliable measures have been developed to appraise knowledge of ethical issues (Braunschweiger and Goodman, 2007; DuBois et al., 2008; Heitman and Bulger, 2006). What should be recognized here, however, is that the validity of knowledge measures depends on systematic sampling of the domain under consideration. In the case of RCR training evaluation, this domain may reflect ethical knowledge in general, ethical knowledge applying to a particular field, or ethical knowledge specifically provided in training. These differing frameworks for generating knowledge items result in differences in the generality of the conclusions flowing from evaluation studies. Moreover, it should be recognized that possessing knowledge does not ensure that this knowledge is actually applied in making ethical decisions.

Knowledge is often of interest because it provides a basis for formulating *mental models*. Although less commonly employed than performance or knowledge measures, mental model measures have been employed in evaluation of RCR programs. Broadly speaking, assessments of mental models are based on a direct or an indirect approach. In the direct approach, people are presented with an ethical vignette and a list of concepts that might be used to understand this vignette. They are asked to indicate linkages among these concepts with scores being based on the similarity of their concept linkages to the concept linkages of ethical experts with regard to this scenario. Brock et al. (2008) provide an illustration of this type of evaluation measure in the context of ethics in the physical sciences and engineering. In the indirect approach, mental model quality is assessed through recognition of the significance of ethical issues or moral sensitivity. Here people are presented with multiple short scenarios where attributes of the scenario relevant to moral sensitivity (e.g., number effected, size of effects, emotional salience) are manipulated. People are asked to indicate which scenarios are most significant. An illustration of this type of measure in the assessment of scientific ethics has been provided by Clarkeburn (2002). Regardless of the approach applied, however, generalization from mental model measures to actual ethical conduct is a matter of inference.

Performance, knowledge, and mental model measures reflect changes in individual capacities as a result of RCR training. However, RCR training may also result in changes in attitudes toward ethics, perception of ethical issues, and interactions with coworkers. These attitudinal effects of RCR instruction are often subsumed under the rubric of *climate*. Climate measures ask people to indicate the extent to which they would endorse ethical behaviors—for example, "I think about my contributions to a manuscript before assigning authorship." Development of viable climate measures, of course, requires identification of behaviors marking ethical conduct in a particular workplace. Thus the generality of infer-

ences is limited by work setting. However, valid and reliable measures of ethical climate for scientific work have been developed (Anderson, 2010; Thrush et al., 2007). Moreover, use of climate measures may prove attractive because, with appropriate aggregation procedures, they may allow assessment of the effects of instruction of teams, departments, or institutions.

Many RCR courses ask students to produce certain *products* as part of instructional exercises. For example, Mumford and coworkers' (2008) instructional program asks students to provide written self-reflections at the end of training. These written self-reflections can be coded by judges for attributes such as ethical awareness, self-objectivity, and appraisal of ethical ambiguities. Similarly, judges may observe students' participation in discussions to assess attributes such as engagement in ethical issues, identification of critical features of the issue, and production of viable solution strategies. Product-based evaluation of educational interventions, often described as portfolio assessments, have gained widespread acceptance in recent years (Reynolds et al., 2009; Slater, 1996). However, use of these techniques is contingent on the availability of a trained cadre of judges who have time to devote to the evaluation process, both requirements that limit widespread application of this evaluation technique in RCR training (Stecher, 1998). Moreover, the nature of these measures makes assessment of change difficult unless parallel exercises have been developed for early-cycle and late-cycle instruction.

An alternative to product assessments is to seek appraisals of instructional content from students. These *reaction* measures are widely applied in evaluation of RCR instruction. A typical reaction question might ask how much did you learn from this course or how much did you enjoy this "case exercise." Because students are being trained, their expertise for appraising instruction is open to question. As a result, reaction measures are not often employed in formal course appraisal. By the same token such measures can indicate engagement in the instructional course. Moreover, students often appear more accurate in their appraisal of specific training exercises. As a result, reaction measures are often used to appraise the effectiveness of instructional techniques and revise instructional approaches. However, the very nature of reaction measures, like production measures, makes it difficult to evaluate change as a result of interventions.

A final approach that might be used to appraise the effectiveness of RCR instruction may be found in *organizational outcomes*. For example, a drop in the number of ethics cases brought to university officials following introduction of an RCR program represents one such measure. Alternatively, student referral of ethical breeches for investigation might be used as another organizational outcome measure. Because of their objective nature, organizational evaluations are often considered to provide rather compelling evidence for the effectiveness of an RCR educational program (Council of Graduate Schools, 2012). By the same token, these measures are often subject to a variety of contaminating variables—the effects of which must be controlled in evaluation. Moreover, orga-

nizational outcomes represent a distal or downstream outcome, and so effects of RCR instruction may take some time, multiple years, to be capable of being observed. As a result of these considerations, as well as access and record-keeping issues, organizational outcomes have not commonly been used in evaluating the effectiveness of RCR training.

## EVALUATION OF RCR TRAINING

### Meta-Analyses

Although a variety of measures are available for evaluation of RCR training, systematic evaluation of the effectiveness of instruction has been sporadic. Some programs have been evaluated while others have not. Nonetheless, enough programs have been evaluated (e.g., Clarkeburn et al., 2002; Gual, 1987; Self et al., 1993) to permit application of meta-analytic procedures (Arthur et al., 2001; Hunter and Schmidt, 2004) in appraising the effectiveness of RCR instruction. In meta-analyses, the cumulative effects observed as a result of an intervention, or measure, across studies are assessed. As a result, meta-analyses provide a basis for evaluating the general effectiveness of current RCR training.

Antes et al. (2009) conducted a meta-analytic study intended to assess the effectiveness of RCR training. They identified 26 prior studies where the effectiveness of ethics instruction in the sciences had been conducted. These studies included 3,041 individuals, primarily individuals in doctoral programs, who received instruction. The effectiveness of instruction was typically appraised by examining changes in ethical decision making, a performance measure, using the Defining Issues Test (Rest, 1988) or Kohlberg's (1976) moral development measure. However, some studies used field-specific ethical decision-making measures. The effects of instruction across studies was assessed using Cohen's Δ—an unstandardized estimate of effect size. In addition, judges content-coded each study with respect to design (e.g., pre-post, pre-post plus controls), participant characteristics (e.g., educational level, field, gender), instructional content (e.g., type of objectives, coverage of ethical standards), and instructional method (e.g., length of instruction, amount of practice, use of multiple practice activities).

The overall Cohen's Δ obtained in this meta-analysis was .42. A Cohen's Δ of .42 indicates that the effectiveness of instruction has weak, albeit beneficial, effects given current standards holding that Cohen's Δ below .40 indicates little effect, between .40 and .80 some effect, and above .80 sizeable effects. However, studies using stronger designs, and stronger instructional programs, typically produced larger effects. More specifically, the most effective programs were longer (more than 9 hours), focused on real-world ethics cases, distributed practice exercises, used multiple types of practice exercises, and had substantial

instructor–student interaction. In courses meeting these criteria, Cohen's Δs were in the .50 to .70 range.

These findings indicate that with respect to performance, RCR training is marginally effective. However, the effectiveness of this instruction increases when more effective educational practices focusing on active application of ethical principles to real-world problems are incorporated in instruction. By the same token it should be recognized that these studies have focused on performance criteria. Although use of performance criteria is desirable, it should be recognized that these findings do not speak to other criteria, knowledge, climate, and organizational outcomes that might be used to evaluate the effectiveness of RCR training.

As is the case in any meta-analytic study the obtained findings depend on the nature of the available archival data. In the Antes et al. (2009) study many of the studies examined had been based on funding from external sources. As a result, questions arise if similar effects would be observed if RCR instruction is provided routinely as opposed to "special" funded initiatives. Moreover, the measures used to assess performance in many of these studies were based on general, non-field-specific, measures of ethical decision making (Rest, 1988).

To address these issues, an additional study was conducted by Antes et al., (2010). The measure used to appraise performance in this study was a field-specific measure of ethical decision making developed by Mumford et al. (2006). On this measure people are presented with an ethical vignette applying in their field—measures having been developed for the following fields: (1) health sciences, (2) social sciences, (3) biological sciences, (4) physical sciences and engineering, (5) the humanities, and (6) performance fields (e.g., arts, architecture). After reading through a vignette, people are presented with a series of three or four events arising in this scenario. For each event they are asked to select two of the 8 to 12 potential responses to the event presented where responses vary with respect to ethical content in terms of data management (e.g., data trimming), study conduct (e.g., informed consent), professional practices (e.g., maintaining objectivity), and business practices (e.g., conflicts of interest). Studies by Helton-Fauth et al. (2003) and Stenmark et al. (2011) have provided evidence for the relevance of these dimensions across fields.

More centrally, a number of studies have provided evidence for the construct validity of these measures of ethical decision-making performance (Antes et al., 2007; Mumford, Connelly, et al., 2009; Mumford et al., 2006, 2007, 2010; Mumford, Waples et al., 2009). Broadly speaking, the findings obtained in these studies indicate: (1) the pre-post versions of these measure evidence adequate reliability (reliability coefficients above .70), (2) scores on these measures are not influenced by social desirability and acquiescence, (3) ethical decision making as assessed by these measures is negatively related to cynicism and narcissism, (4) scores on these measures are positively related to punitive actions taken in response to ethical breeches, (5) scores on these measures are positively related

to creative problem solving, (6) scores on these measures are negatively related to perceptions of interpersonal conflict in the work environment, and (7) scores on the measure are negatively related to exposure to unethical practices in their day-to-day work. Thus a compelling body of evidence is available for the construct validity of Mumford and colleagues' (2006) measures of ethical decision making.

Antes et al. (2010) administered the health, biological, and social sciences measures in 21 RCR courses providing training for 173 doctoral students at major research universities. These measures were administered in a pre-post design and the effectiveness of RCR instruction was assessed. It was found that in these courses trivial, nonexistent, effects of instruction on ethical decision-making performance were observed—Cohen's $\Delta$ = -.08. Moreover, analysis of responses to these measures suggested these weak effects might be due to induction of self-protection and self-enhancement (e.g., I've been trained and am therefore ethical) as a result of RCR training. Thus although RCR training has value, its value may not always be maintained when instruction becomes institutionalized. This finding points to the importance of ongoing evaluation of the effectiveness of RCR instruction.

### Ongoing Evaluation of an Exemplar Program

The findings obtained in the Antes et al. (2009) study provided a basis for developing an instructional program on professional ethics and the responsible conduct of research. This instructional program is given to all students receiving stipends, either research stipends or teaching assistant stipends at the University of Oklahoma—some 600 students annually. Ongoing evaluation was expressly "built into" the design of this RCR program with the program being structured in such a way that new instructional initiatives could also be evaluated.

Mumford et al. (2008) and Kligyte et al. (2008) provide a description of this instructional program. The substantive basis for this instructional program was that ethical decision making in real-world settings depends on sense making (Sonenshein, 2007) or understanding the consequences of actions for various stakeholders. Within this sense-making framework it is held that ethical guidelines, prior professional experience, professional goals, and affect all influence peoples' decisions (Mumford et al., 2008) along with the strategies people employ in working with this information to make decisions—strategies such as framing situations in terms of ethical implications, analyzing motivations, questioning judgments, regulating emotions, forecasting downstream implications of actions, and considering the effects of actions on relevant stakeholders (Thiel et al., 2012).

Instruction in sensemaking is provided over 2 days, through 10 blocks of instruction, in a peer-based cooperative learning framework. Instructors in this face-to-face instruction are trained, senior, doctoral students. The instruction occurs in the context of cases and exercises (e.g., role plays) intended to illustrate real-world application of key principles being covered in a given block

of instruction. The instructional program consists of 10 blocks of instruction examining: (1) ethical research guidelines, (2) complexity in ethical decision making, (3) personal biases in ethical decision making, (4) problems encountered in ethical decision making, (5) ethical decision-making sense-making strategies, (6) field-specific differences in applying decision-making strategies, (7) sense making in ethical decision making, (8) complex field differences, (9) understanding the perspectives of different stakeholders, and (10) applying knowledge gained in training.

Prior to instruction, participants are asked to complete the pre-test ethical decision-making measure applicable to their field (e.g., biological sciences, physical sciences, and engineering) and after training they are asked to complete the post-test measure. These pre-post measures were drawn from the earlier work of Mumford et al. (2006). Pre-post comparisons are used to assess change in ethical decision making applying either a normative scoring model or an Angoff model—where changes in pass rates are assessed with respect to an expert's definition of minimally acceptable ethical decisions. In addition, after each day of instruction, participants' reactions to instruction are assessed with respect to appraisals, or a seven-point scale, of the value of the exercises presented in each block of instruction. Both the performance and reaction measures are obtained in each class, and relevant evaluation data are examined biyearly.

Evaluation of the impact of this instruction on ethical decision making has been described by Mumford et al. (2008) and Kligyte et al. (2008). In these studies the normative scoring format was used in assessing pre-post change. They found this instruction resulted in Cohen's $\Delta$ between .49 and 1.82 across decisions involving data management, study conduct, professional practices, and business practices. The average effect size was .91. When scored using the Angoff method, reflecting changes from a priori pass rates, the resulting Cohen's $\Delta$s range between .70 and 2.4, producing an average effect size estimate of 1.4. The larger effects obtained for Angoff scores are the result of range restriction suppressing variance when a normative scoring method is employed. Moreover, these effects have been maintained over a 5-year period where instructors have been rotated in and out. Thus this sense-making instruction apparently results in sizeable effects on ethical decision making, a performance measure, with these effects being maintained over time—in other words they are not instructor or class specific.

A second key piece of evaluation evidence is provided by an alternative scoring of the ethical decision-making measure. Responses on the ethical decision-making measure also allow for scoring of the application of key sense-making strategies (e.g., recognizing circumstances, anticipating consequences, considering others' perspectives). Scoring for use of these strategies is noteworthy because the instructional program is intended to encourage the use of more effective strategies in ethical decision making. In fact, the findings obtained in the Kligyte

et al. (2008) and Mumford et al. (2008) studies indicate sizeable gains, average Cohen's $\Delta = .7$, in application of viable sense-making strategies.

Of course, the data gathered on these sensemaking strategies is embedded in the ethical decision-making measure. As a result, a series of independent experimental investigations were conducted by Antes et al. (2012), Stenmark et al. (2010, 2011), Brown et al. (2011), Caughron et al. (2011), Martin et al. (2011), and Thiel et al. (2011). In these studies manipulations were made to induce application of more effective sense-making strategies—for example, induction of an analytical mindset or induction of self-reflection of prior experience. Participants in these studies were then assessed for performance in strategy execution and ethical decision making. The findings obtained in these studies indicated that effective application of these sense-making strategies contributed to more effective ethical decision making. Thus these studies served to provide evidence for the meaningfulness of the decision-making strategies being trained. Moreover, these studies illustrate the value of incorporating independent studies in evaluation programs expressly intended to appraise the merits of substantial assumptions underlying development of curriculum and instructional approach.

To assess the impact of this instructional program with respect to mental models, an alternative approach based on experimental methods was employed. In the Brock et al. (2008) study, three groups were identified. One group had been asked to complete the professional ethics education program 6 months earlier. The second group was a cohort of doctoral students who had not received the training. The third group were faculty working in the same field who had not completed training. Participants were presented with four ethical scenarios—one examining ethical issues with respect to data management, study conduct, professional practices, and business practices. Think-aloud protocols were obtained as members of each group worked through these scenarios to arrive at a decision. Subsequently, judges coded these transcripts with respect to 15 dimensions such as goal assessment, perceived threats, information integration, and norm-based framing evident in participants' verbalizations. A pathfinder analysis was used to identify the mental models employed by each group. It was found that the models employed by faculty and untrained doctoral students stressed environmental monitoring in relation to experience and personal values to reach ethical decisions. In contrast, the models used by trained doctoral students stressed problem appraisal from the perspective of others and solution appraisal (forecasting) along with contingency planning. Thus ethics education apparently resulted in the acquisition of stronger mental models—stronger mental models which were maintained over a 6-month period and were evident on transfer tasks.

In addition to improvements in ethical decision making and ethical decision-making strategies, both performance measures, and improvements in the mental models used to understand ethical problems—improvements maintained over a 6-month period on transfer tasks—evaluation of this professional ethical instructional program has also considered student reactions. These reaction measures

are collected at the end of each day of instruction. On these measures students are asked to rate, on a seven-point scale, how favorably they reacted to the cases, exercises, and discussion embedded in each block of instruction. Kligyte et al. (2008) have shown these reaction measures evidence adequate reliability—reliability coefficients above .70. More centrally, students generally expressed positive appraisals of the cases, exercises, and discussions occurring in each block of instruction with mean ratings ranging between 5.0 and 6.5 on a seven-point scale. Again, these positive reactions have been maintained over 5 years and across multiple instructors. Although low student appraisals would have led to changes in instructional content, the positive nature of the students' reactions, in light of findings bearing on performance and mental models, did not indicate the need to make significant revisions in instructional content. This observation is of some importance because it points to the need to appraise reaction data in the light of other data bearing on program effectiveness.

The final method used to appraise the effectiveness of this instructional program has been an ongoing analysis of critical incidents occurring at the organizational level involving incidents of ethical misconduct. When the program was established, access to organizational responses to ethical breaches was obtained through the office of the graduate dean. These metrics are appraised using qualitative methods including discussion of ethical issues arising and responses to these issues in a biannual meeting of the graduate dean and director of the ethics education program. Three general organizational outcomes have been observed following implementation of this ethics education program. First, the number of "false" complaints of ethical misconduct presented to the graduate dean has declined. Second, issues involving significant incidents of ethical misconduct are reported to the graduate dean more quickly and the institution has responded in a more timely fashion to these incidents of misconduct. Third, the people reporting these incidents of misconduct are doctoral students who have completed the professional ethics/responsible conduct of research education program.

Taken as a whole, the sense-making RCR education program appears effective with respect to performance, mental models, reactions, and organizational outcome evaluation criteria. Although some criteria, for example, climate and knowledge, have not been examined, the pattern of evidence suggests the program *may* also be beneficial, or at least not disruptive, with regard to these attributes of RCR outcomes. Moreover, the beneficial effects of sense-making instruction are apparently maintained over time and on transfer tasks. As a result, these measures are used in routine evaluation of both the overall instructional program and evaluation of the effectiveness of individual instructors, with poor instruction resulting in remedial training for instructors or dismissal of ineffective instructors. Thus the evaluation data are actively used in day-to-day administration of this RCR ethics education program.

## Evaluation and RCR Instruction

In instructional systems, including RCR training, evaluation is commonly viewed in a distinct way. We assume once an evaluation study, or set of evaluation studies, has been conducted, and the findings are positive, no further evaluation is necessary. However, as noted above, evaluation should be an ongoing process providing data needed for day-to-day management of the instructional system. More centrally, instructional systems can be created that permit evaluation of new instructional approaches. The data provided by such initiatives, at least potentially, allows for the ongoing, progressive refinement of instruction including RCR training programs.

A series of studies conducted by Harkrider et al. (2012, 2013); Johnson et al. (2012), Peacock et al. (2013), and Thiel et al. (2013) provide illustrations of the use of evaluation data in continuous improvement in RCR programs. The basis for all these studies was the sense-making RCR training program developed by Mumford et al. (2008). As noted earlier, this program consisted of 10 blocks of instruction. An additional, one-and-a-half-hour block was added at the beginning of the second day of instruction. This block of instruction focused on the implications of ethical cases. All these studies examined merits of different approaches to the presentation of case material in RCR instruction.

All these studies presented one or two cases describing complex ethical issues where breaches in ethical conduct occurred. Experiments were then conducted by varying the aspects of case content presented on how participants were instructed to work with case content. For example, in the Thiel et al. (2013) study, case content was manipulated to stress, or not mention, emotional consequences of the events described in the case for key stakeholders. In the Peacock et al. (2013) study, participants either were, or were not, asked to consider the effects of alternative outcomes of the case scenario.

In all studies, four evaluation measures were used to assess the effects of these manipulations on ethics. The first evaluation measure, a knowledge measure, completed at the end of this block of instruction, examined retention of key information in the cases presented. The second, a transfer task, presented again at the end of the block of instruction, asked participants to answer questions bearing on another ethical case. These open-ended responses were coded by four trained judges, judges evidencing adequate agreement, for decision ethicality, recognition of critical causes, recognition of critical constraints, and forecast quality. Third, at the end of the instructional day, participants' reactions to the instruction they received were obtained. Fourth, and finally, at the end of instruction, participants were asked to complete the Mumford et al. (2006) measure of ethical decision making which also provided measures of the ethical decision-making strategies people employed.

The findings obtained in these studies, all findings based on the evaluation measures described above, have been informative as to how case material should be used in ethics education. For example, the findings obtained by Thiel et al.

(2013) indicated that tagging stakeholders' emotional reactions in cases results in better knowledge acquisition, better ethical decision making and strategy application on the transfer task, and better ethical decision making at the end of instruction. The Peacock et al. (2013) study indicated that presenting alternative outcome scenarios to the case reduced knowledge acquisition, use of ethical decision-making strategies, and ethical decision making at the end of training. These findings are noteworthy because they suggest that overcomplication of case material may diminish knowledge acquisition and subsequent ethical decision making. In the Harkrider et al. (2012) study, it was found that when cases were linked to codes of conduct and forecasts based on the case were made with respect to codes of conduct, knowledge acquisition, ethical decision making and strategies for ethical decision making on the transfer task, and end-of-instruction ethical decision making all improved.

The findings obtained in these studies, of course, illustrate not only the use of knowledge measures on the evaluation of RCR instruction, they also illustrate how systematic evaluation programs can be "built into" ongoing programs of instruction. More specifically, blocks of instruction can be isolated where "field" experiments can be conducted. The results flowing from these studies, in turn, provide a basis for revision of other curriculum content while adding to the knowledge of how RCR training should be conducted. Thus RCR evaluation should be viewed as a dynamic, ongoing process with our understanding of the requirements for effective RCR education improving over time.

## CONCLUSIONS

The present effort broaches an important and basic question. Does RCR education work? Any attempt to answer this question must bear in mind the issue "work with respect to what." RCR education programs might be evaluated with respect to changes in ethical decision-making performance, knowledge of ethics, the mental models people employ to understand ethical issues, perceptions of ethical climate, the products people produce, reactions to instruction, and organizational outcomes. Prior evaluation efforts have focused primarily, almost exclusively, on ethical decision-making performance (Antes et al., 2009).

Bearing in mind that the available data do not speak to many of the evaluation criteria that might be applied, the findings obtained by Antes et al. (2009) in their meta-analysis indicated that RCR training has only weak, marginal, effects on ethical decision making. Moreover, the findings obtained by Antes et al. (2010) indicate that as RCR training is executed in a day-to-day fashion such instruction may have no effect on ethical decision making when valid, reliable measures of ethical decision making are employed. Given the fact that the intent of most RCR training is to improve performance, the findings emerging from these studies are troublesome.

By the same token, the Antes et al. (2008) study did not indicate that all

programs fail. The RCR programs that proved especially effective were lengthy, in-depth courses that presented multiple real-world ethics cases where students were encouraged to work through these cases, or exercises, in an active, cooperative, instructional format. These principles provided the background information underlying development of Mumford and colleagues' (2008) sense-making training. The findings obtained in evaluation of this RCR/professional ethics program indicate that it resulted in substantial gains in students' ethical decision-making performance, gains in the viability of students' mental models for understanding ethical issues, and gains that were maintained over time and across cohorts. Moreover, students reacted positively to this instruction, and positive changes in organizational outcomes were observed. Thus well-developed RCR training programs can work and work with respect to multiple measures of program performance.

Although other research supports the key principles underlying development of this program for ethics education (Thiel et al., 2012), it is also true that this program is not the only potentially viable approach that might be taken to ethics instruction. Other substantive models of ethics exist (Haidt, 2001) and some of those alternative models may prove more appropriate when instruction in RCR or professional ethics has other goals (e.g., Braunschweiger and Goodman, 2007)—for example, improving mastery of ethical guidelines as opposed to improving ethical decision making. Nonetheless, the evaluation data gathered for this program are noteworthy not only because they indicate that RCR training can work but that viable RCR training is most likely to be developed when courses are designed to take into account the findings obtained in earlier evaluation studies. Moreover, evaluation may be embedded in instructional programs as an ongoing element of instruction (e.g., Thiel et al. 2013) thereby providing a stronger, richer, basis for evaluating key elements of ethics instruction such as the use of cases. One hopes that the present effort will provide an impetus for ongoing, systematic, and multifaceted evaluation of RCR training. It is only through the findings of these evaluation studies that we will be able to formulate RCR training programs that have real effects on the ethical conduct of our scientists and the organizations in which they work.

## REFERENCES FOR APPENDIX C

Aguinis, H., and Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology*, *60,* 451-474.

Anderson, M. S. (2010). *Creating the Ethical Academy: A Systems Approach to Understanding Misconduct and Empowering Change*. New York: Routledge.

Antes, A. L., Brown, R. P., Murphy, S. T., Waples, E. P., Mumford, M. D., Connelly, S., and Devenport, L. D. (2007). Personality and ethical decision-making in research: The role of perceptions of self and others. *Journal of Empirical Research on Human Research Ethics, 2*, 15-34.

Antes, A. L., Murphy, S. T., Waples, E. P., Mumford, M. D., Brown, R. P., Connelly, S., and Devenport, L. D. (2009). A meta-analysis of ethics instruction effectiveness in the sciences. *Ethics & Behavior*, *19*(5), 379-402.

Antes, A. L., Wang, X., Mumford, M. D., Brown, R. P., Connelly, M. S., and Devenport, L. D. (2010). Evaluating the effects that existing instruction of responsible conduct of research has on ethical decision making. *Academic Medicine, 85*, 514-526.

Antes, A. L., Thiel, C. E., Martin, L. E., Stenmark, C. K., Connelly, S., Devenport, L. D., and Mumford, M. D. (2012). Applying cases to solve ethical problems: The significance of positive and process-oriented reflection. *Ethics & Behavior*, *22*(2), 113-130.

Arthur, W., Bennett, W., and Huffcutt, A. (2001). *Conducting Meta-analysis Using SAS*. Mahwah, NJ: Lawrence Erlbaum Associates.

Baldwin, T. T., and Ford, J. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, *41*(1), 63-105.

Bebeau, M. J., and Thoma, J. (1994). The impact of a dental ethics curriculum on moral reasoning. *Journal of Dental Education, 58*, 684-692.

Braunschweiger, P., and Goodman, K. W. (2007). The CITI program: An international online resource for education in human subjects protection and the responsible conduct of research. *Academic Medicine*, *82*(9), 861-864.

Brock, M. E., Vert, A., Kligyte, V., Waples, E. P., Sevier, S. T., and Mumford, M. D. (2008). Mental models: An alternative evaluation of a sensemaking approach to ethics instruction. *Science and Engineering Ethics, 14*, 449-472.

Brown, R. P., Tamborski, M., Wang, X., Barnes, C. D., Mumford, M. D., Connelly, S., and Devenport, L. D. (2011). Moral credentialing and the rationalization of misconduct. *Ethics & Behavior*, *21*(1), 1-12.

Cascio, W., and Aguinis, H. (2004). *Applied Psychology in Human Resource Management*. New York: Pearson.

Caughron, J. J., Antes, A. L., Stenmark, C. K., Thiel, C. E., Wang, X., and Mumford, M. D. (2011). Sensemaking strategies for ethical decision making. *Ethics & Behavior*, *21*(5), 351-366.

Clarkeburn, H. (2002). A test for ethical sensitivity in science. *Journal of Moral Education*, *31*(4), 439-453.

Clarkeburn, H., Downie, J. R., and Matthew, B. (2002). Impact of an ethics programme in a life sciences curriculum. *Teaching in Higher Education, 7*, 65-79.

Colquitt, J. A., LePine, J. A., and Noe, R. A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, *85*(5), 678-707.

Cook, T. D., and Campbell, D. T. (1979) *Quasi-Experimentation: Design and Analysis for Field Settings*. Chicago: Rand-McNally.

Council of Graduate Schools. (2012). *Research and Scholarly Integrity in Graduate Education*. Washington, DC: Council of Graduate Schools.

Drake, M. J., Griffin, P. M., Kirkman, R., and Swann, J. L. (2005). Engineering ethical curricula: Assessment and comparison of two approaches. *Journal of Engineering Education, 94*, 223-232.

DuBois, J. M., Dueker, J. M. (2009). Teaching and assessing the responsible conduct of research: A Delphi Consensus Panel Report. *Journal of Research Administration, 60*, 49-70.

DuBois, J. M., Dueker, J. M., Anderson, E. E., and Campbell, J. (2008). The development and assessment of an NIH-funded research ethics training program. *Academic Medicine*, *83*(6), 596-603.

Fleishman, E. A., and Mumford, M. D. (1989). Individual attributes and training performance. In *Training and Development in Organizations* (pp. 183-255). San Francisco, CA: Jossey-Bass.

Gawthrop, J. C., and Uhlemann, M. R. (1992). Effects of the problem-solving approach in ethics training. *Professional Psychology: Research and Practice*, *23*, 38-42.

Goldstein, I. L. (1986). *Training in Organizations: Needs Assessment, Development, and Evaluation*. Monterey, CA: Brooks/Cole.

Gottman, T. M. (1995). *The Analysis of Change*. Hillsdale, NJ: Erlbaum.

Gual, A. L. (1987). The effect of a course in nursing ethics on the relationship between ethical choice and ethical action in baccalaureate nursing students. *Journal of Nursing Education, 26*, 113-117.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review*, *108*(4), 814-834.

Harkrider, L. N., MacDougall, A., Bagdasarov, Z., Johnson, J. F., Thiel, C. E., Mumford, M. D., Connelly, M. S., and Devenport, L. D. (2013). Structuring case-based ethics training: How comparing cases and structured prompts influence training effectiveness. *Ethics and Behavior*.

Harkrider, L. N., Thiel, C. E., Bagdasarov, Z., Mumford, M. D., Johnson, J. F., Connelly, S., and Devenport, L. D. (2012). Improving case-based ethics training with codes of conduct and forecasting content. *Ethics & Behavior*, *22*(4), 258-280.

Heitman, E., and Bulger, R. E. (2006). Assessing the educational literature in the responsible conduct of research for core content. *Accountability in Research, 12,* 207-224.

Helton-Fauth, W., Gaddis, B., Scott, G., Mumford, M. D., Devenport, L. D., Connelly, M. S., and Brown, R. P. (2003). A new approach to assessing ethical conduct in scientific work. *Accountability in Research, 10*, 205-228.

Hunter, J. E., and Schmidt, F. L. (2004). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Newberry Park, CA: Sage.

Institute of Medicine and National Research Council. (2002). *Integrity in Scientific Research*. Washington, DC: National Academies Press.

Johnson, J. F., Thiel, C. E., Bagdasarov, Z., Connelly, M. S., Harkrider, L. N., Devenport, L. D., and Mumford, M. D. (2012). Case-based ethics education: The impact of cause complexity and outcome favorability of ethicality. *Journal of Empirical Research on Human Ethics, 7*, 63-77.

Kirkpatrick, D. L. (1978). Evaluating in-house training programs. *Training and Development Journal, 32*, 6-9.

Kligyte, V., Marcy, R. T., Waples, E. P., Sevier, S. T., Godfrey, E. S., Mumford, M. D., and Hougen, D. F. (2008). Application of a sensemaking approach to ethics training for physical sciences and engineering. *Science and Engineering Ethics, 14*(2), 251-278.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In Lickman (Ed.) *Moral Development and Behavior: Theory, Research, and Social Issues* (pp. 12-33). New York: Holt, Reinhart, & Winston.

Major-Kincade, T. L., Tyson, J. E., and Kennedy, K. A. (2001). Training pediatric house staff in evidence-based ethics: an exploratory controlled trial. *Journal of Perinatology, 21*, 161-166.

Martin, L. E., Stenmark, C. K., Thiel, C. E., Antes, A. L., Mumford, M. D., Connelly, S., and Devenport, L. D. (2011). The influence of temporal orientation and affective frame on use of ethical decision-making strategies. *Ethics & Behavior*, *21*(2), 127-146.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741-749.

Motowidlo, S. J., Dunnette, M. D., and Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal Of Applied Psychology*, *75*(6), 640-647.

Mumford, M. D., Connelly, S., Brown, R. P., Murphy, S. T., Hill, J. H., Antes, A. L., Waples, E. P., and Devenport, L. D. (2008). Sensemaking approach to ethics training for scientists: Preliminary evidence of training effectiveness. *Ethics and Behavior, 18*, 315-339.

Mumford, M. D., Connelly, S., Murphy, S. T., Devenport, L. D., Antes, A. L., Brown, R. P., and Waples, E. P. (2009). Field and experience influences on ethical decision making in the sciences. *Ethics & Behavior*, *19*(4), 263-289.

Mumford, M. D., Devenport, L. D., Brown, R. P., Connelly, S., Murphy, S. T., Hill, J. H., and Antes, A. L. (2006). Validation of Ethical Decision Making Measures: Evidence for a New Set of Measures. *Ethics & Behavior*, *16*(4), 319-345.

Mumford, M. D., Murphy, S. T., Connelly, S., Hill, J. H., Antes, A. L., Brown, R. P., and Devenport, L. D. (2007). Environmental influences on ethical decision making: Climate and environmental predictors of research integrity. *Ethics & Behavior*, *17*(4), 337-366.

Mumford, M. D., Waples, E. P., Antes, A. L., Brown, R. P., Connelly, S., Murphy, S. T., and Devenport, L. D. (2010). Creativity and ethics: The relationship of creative and ethical problem-solving. *Creativity Research Journal*, *22*(1), 74-89.

Mumford, M. D., Waples, E. P., Antes, A. L., Murphy, S. T., Connelly, S., Brown, R. P., and Devenport, L. D. (2009). Exposure to unethical career events: Effects on decision making, climate, and socialization. *Ethics & Behavior*, *19*(5), 351-378.

National Academy of Engineering. (2009). *Ethics Education and Scientific and Engineering Research*. Washington, DC: The National Academies Press.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (1992). *Responsible Science: Ensuring Integrity of the Research Process*. Washington, DC: National Academy Press.

Peacock, H. J., Harkrider, L. N., Bagdasarov, Z., Connelly, M. S., Johnson, J. F., Thiel, C. E., MacDougall, A. E., Mumford, M. D., and Devenport, L. D. (2013). Effects of alternative scenarios and structured outcome evaluation on case-based ethics instruction. *Science and Engineering Ethics, 19*(3), 1283-1303.

Rest, J. R. (1988). *DIT Manual: Manual for the Defining Issues Test*. Saint Paul, MN: Center for the Study of Ethical Development.

Reynolds, C. R., Livingston, R. B., Willson, V., and Willson, V. L. (2006). *Measurement and Assessment in Education*. Boston, MA: Allyn & Bacon/Pearson Education.

Self, D. J., Schrader, D. E., Baldwin, D. C., and Wolinsky, F. D. (1993). The moral development of medical students: A pilot study of the possible influence of medical education. *Medical Education*, *27*(1), 26-34.

Shadish, W. R., Cook, T. D., and Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton, Mifflin and Company.

Slater, T. F. (1996). Portfolio assessment strategies for grading first-year university physics students. *Physics Education, 31*, 329-338.

Sonenshein, S. (2007). The role of construction, intuition, and justification in responding to ethical issues at work: The sensemaking-intuition model. *Academy of Management Review*, *32*(4), 1022-1040.

Stecher, B. (1998). The local benefits and burdens of large-scale portfolio assessments. *Assessment in Education, 5*, 335-351.

Steneck, N. H., and Bulger, R. (2007). The history, purpose, and future of instruction in the responsible conduct of research. *Academic Medicine*, *82*(9), 829-834.

Stenmark, C. K., Antes, A. L., Thiel, C. E., Caughron, J. J., Wang, X., and Mumford, M. D. (2011). Consequences identification in forecasting and ethical decision-making. *Journal of Empirical Research on Human Research Ethics*, *6*(1), 25-32.

Stenmark, C. K., Antes, A. L., Wang, X., Caughron, J. J., Thiel, C. E., and Mumford, M. D. (2010). Strategies in forecasting outcomes in ethical decision-making: Identifying and analyzing the causes of the problem. *Ethics & Behavior*, *20*(2), 110-127.

Thiel, C., Connelly, S., and Griffith, J. (2011). The influence of anger on ethical decision making: Comparison of a primary and secondary appraisal. *Ethics & Behavior*, *21*(5), 380-403.

Thiel, C. E., Bagdasarov, Z., Harkrider, L., Johnson, J. F., and Mumford, M. D. (2012). Leader ethical decision-making in organizations: Strategies for sensemaking. *Journal of Business Ethics*, *107*(1), 49-64.

Thiel, C. E., Connelly, M. S., Harkrider, L. N., Devenport, L. D., Bagdasarov, Z., Johnson, J. F., and Mumford, M. D. (2013). Case-based knowledge and ethics education: Improving learning and transfer through emotionally rich cases. *Science and Engineering Ethics, 19*(1), 265-286.

Thrush, C. R., Vander Putten, J., Rapp, C., Pearson, L., Berry, K., and O'Sullivan, P. S. (2007). Content validation of the Organizational Climate for Research Integrity (OCRI) survey. *Journal of Empirical Research on Human Research Ethics*, *2*(4), 35-52.

Yammarino, F. J., Dionne, S. D., Chun, J., and Dansereau, F. (2005). Leadership and levels of analysis: A state-of-the-science review. *Leadership Quarterly*, *16*(6), 879-919.

# Appendix D

# Detailed Case Histories

The following five detailed case histories of specific cases of actual and alleged research misconduct are included in an appendix to raise key issues and impart lessons that underlie the committee's findings and recommendations without breaking up the flow of the report. In several cases, including the translational omics case at Duke University and the Goodwin case at the University of Wisconsin, the committee heard directly from some of those involved.

The case histories differ in length in order to devote sufficient explanation to the issues involved in each case. For example, the translational omics case at Duke University unfolded over several years and involved multiple complex issues, making a lengthier discussion necessary. Issues covered in the cases include individual and institutional conflicts of interest, data falsification and fabrication, whistleblower retaliation and protection, insufficient or abusive mentoring, ghostwriting, authorship roles, institutional and administrator responsibilities, journal responsibilities, implementation of the federal government's research misconduct policy, and the costs and impacts of research misconduct.

*Some cases mentioned in the report are not included in the appendix because the shorter descriptions already sufficed to illustrate the issues being described.*

## THE WAKEFIELD MMR-AUTISM CASE

***Synopsis and Rationale for Inclusion:*** *An undisclosed conflict of interest between a principal investigator and the entity funding their research can have far-reaching effects beyond the scope of the research study. In the MMR-autism case, Andrew Wakefield had undisclosed monetary conflicts of interest and was found to have violated human subjects protection rules in research underlying*

*an article published in the* Lancet *(UK GMC, 2010; Triggle, 2010).*[1] *In the opinion of the* British Medical Journal*, Wakefield also falsified data (Godlee et al., 2011). A formal retraction did not occur for over a decade, allowing ample time for the purported findings to become an important support for the anti-vaccine movement. This case not only confronts the issue of conflicts of interest but also weaknesses in institutional research governance, coauthor responsibility, and journal responsibility.*

In 1998, Andrew Wakefield published a paper in *The Lancet* claiming that he had found a link between the measles, mumps, and rubella (MMR) 3-in-1 vaccine and regressive autism, as well as a bowel disorder, using a sample of 12 children. Within a year, an article with a sample of 498 children rebutted Wakefield's findings, followed by additional rebuttal articles for several years thereafter (Taylor et al., 1999). However, Wakefield's article resonated with anti-vaccine movements in several countries, especially in the United Kingdom and United States, prompting some parents to refrain from vaccinating their children for fear of a connection to autism, contributing to decreased vaccination rates in the United States and United Kingdom and compromising the near success of eradicating these diseases from Western countries.

Six years after the 1998 article was published, 10 of the 12 coauthors retracted the paper's interpretation that the results suggested a possible causal link between the MMR vaccine and autism (Murch et al., 2004). In 2010, based on the UK General Medical Council's (GMC) Fitness to Practice Panel findings, *The Lancet* retracted the full article (*Lancet* Editors, 2010). Both of these retractions were prompted by the investigation by a British journalist, Brian Deer, initially published in the *Sunday Times* in early 2004. Deer exposed that Wakefield had undisclosed financial interest in the research results, reporting that Wakefield had negotiated a contract with a lawyer who hired him to provide evidence against the MMR vaccine to help support a lawsuit against the MMR manufacturing company (Deer, 2011a). Deer reported that Wakefield profited approximately $750,000 USD from the partnership (Deer, 2011a). In addition, Deer stated that Wakefield applied for a patent on his own measles vaccine, from which he was positioned to personally profit (Deer, 2011a). In addition, Deer reported that throughout the study, "Wakefield had repeatedly changed, misreported and misrepresented diagnoses, histories and descriptions of the children, which made it appear that the syndrome had been discovered" (Deer, 2011a). Lastly, Deer reported that the study sample was selectively recruited and not consecutively chosen as Wakefield had reported (Deer, 2011a; Wakefield et al., 1998, retracted). Deer then broadcast his findings on a UK television program, excerpts of which

[1] The United Kingdom General Medical Council's findings of fact from its January 2010 hearing are available in document form. Its verdict finding Wakefield guilty of serious professional misconduct and decision to strike him from the medical register are not available in document form, having been read aloud at a May 2010 hearing, so a news report of this hearing is cited.

were later broadcast in the United States during an NBC *Dateline* investigation on Wakefield.

In addition to Deer's findings, the GMC found that Wakefield had performed unnecessary invasive tests on children that were "against their best interests," was not qualified to perform the tests, did not have the necessary ethics approval to conduct his study, and unethically gathered blood samples by paying children at his son's birthday party for samples (Triggle, 2010; UK GMC, 2010). He was found guilty of more than 30 charges of serious professional misconduct and removed from the UK's medical register (Triggle, 2010; UK GMC, 2010).

Also in 2004 and soon after Deer's investigation, *The Lancet* launched an investigation of the paper. Other than undisclosed parallel funding and ongoing litigation, the *Lancet* reported that their editors did not find evidence of intentional deception or data falsification and so did not retract the paper (Eggertson, 2010). The article remained in the publication until the GMC's findings and subsequent actions in 2010, at which point *The Lancet* editors agreed "several elements of the 1998 paper by Wakefield et al are incorrect, contrary to the findings of an earlier investigation" and fully retracted the paper (*Lancet* Editors, 2010). The journal's editor, Richard Horton, said that "he did not have the evidence to [retract the paper] before the end of the GMC investigation" (Boseley, 2010).

In 2011, Brian Deer produced additional investigative reporting in support of his allegation that Wakefield falsified data, which was published by the British Medical Journal (Deer, 2011b). Deer's work was endorsed by the editors of BMJ (Godlee et al., 2011).

Wakefield denies ever having committed research misconduct; in a press complaint, Wakefield insisted "he never claimed that the children had regressive autism, nor that they were previously normal . . . never misreported or changed any findings in the study, never patented a measles vaccine . . . and he never received huge payments from the lawyer" (Deer, 2011b). Furthermore, he claims to be a victim of conspiracy via a Centers for Disease Control (CDC) cover-up, alleging the "CDC has known for years about an association between the MMR vaccine and autism" (Ziv, 2015). Wakefield's recent basis of this claim is a 2014 article by Brian Hooker published in *Translational Neurodegeneration* in which Hooker reevaluates data collected by the CDC and suggests African American boys who received the MMR vaccine before 24 months and after 36 months of age showed higher risks for autism (Hooker, 2014, retracted). However, the Hooker paper was later retracted because of conflicts of interest and questionable research methods (*Translational Neurodegeneration* Editor and Publisher, 2014).

Following the 2004 investigation, Wakefield moved to the United States, where he is not licensed, but continues to defend the MMR-autism connection. He attempted to sue Deer and the *BMJ* in 2010 for defamation, but the lawsuit was dismissed (Lindell, 2014). Wakefield works out of Austin, Texas, as an anti-vaccine activist, where he has received support from parents of children with autism (Deer, 2014). He directed the documentary *Vaxxed: From Cover-Up to*

*Catastrophe*, which was to have been shown at the 2016 Tribeca Film Festival, but was withdrawn (Goodman, 2016).

In March 2011, the University College London (UCL), which took over the Royal Free Hospital where Wakefield worked at the time, announced intentions to conduct an institutional investigation on Wakefield (Reich, 2011). However, over 1 year later, UCL had not completed the investigation and explained that "given the passage of time, the fact that the majority of the main figures involved no longer work for UCL, and the fact that UCL lacks any legal powers of compulsion," an investigation would not be a worthwhile endeavor for the university (UCL, 2012). Instead, UCL published a paper, *MMR and the Development of a Research Governance Framework in UCL*, detailing revisions made to the university's research governance framework in response to the shortcomings raised by the Wakefield case.

## PAXIL CASE

***Synopsis and Rationale for Inclusion:*** *The Paxil case illustrates issues related to biomedical ghostwriting and unacknowledged conflicts of interest. In this practice, the listed authors of an article reporting on a clinical study may consist solely of prominent academicians, yet unacknowledged industry-supported researchers may have undertaken key tasks associated with the research, including aspects of concept design, subject enrollment, monitoring, data collection and interpretation, and writing the article. In extreme cases, the listed authors may not be able to confirm the integrity of the data or reported results. There have also been several notable cases over the past several decades in which suppression of negative findings or data falsification have been alleged or confirmed in industry-supported studies. Biomedical ghostwriting has been condemned by numerous scientific organizations worldwide.*

Ghostwriting, "the practice whereby individuals make significant contributions to writing a manuscript but are not named as authors," has been condemned as an "example of fraud" and "a disturbing violation of academic integrity standards, which form the basis of scientific reliability" (Bosch and Ross, 2012; Stern and Lemmens, 2011). The practice is not currently equated with plagiarism and so is not within the Office of Research Integrity's (ORI) power to regulate. Bosch and Ross (2012) suggest that ORI include ghostwriting in its definition of research misconduct so that it can be investigated and offenders can be punished under the federal research misconduct policy.

ICMJE (2015) established criteria against which to determine appropriate assignment of biomedical authorship and recommends that those who do not meet all of the criteria only be listed in the acknowledgments sections. COPE (2011) also recommends that specific rules be implemented to prevent ghostwriting, which is explicitly defined as misconduct in their guidelines.

If data are falsified or the reported results are misleading in a clinical study and the listed authors are not able to vouch for the integrity of the data or results, using the study as a basis for treating patients may present serious health and safety risks. If fabricated or falsified results are alleged for privately funded research, institutions are not required to report the investigation results to federal agencies under the federal research misconduct policy.

One example that illustrates these two issues is a 2001 paper overstating the benefits and understating the risks of the Glaxo SmithKline (GSK) drug Paxil in off-label treatment of children (Basken, 2012). Four GSK employees acted as whistleblowers, revealing "improper practices" to the U.S. government, including GSK enticing doctors with vacations and knowingly publishing misreported data (Thomas and Schmidt, 2012). Although the lead authors listed on the paper were respected academics in the field, as part of Glaxo's $3 billion settlement with the federal government, the company admitted that it had hired authors who were not listed as such and that the resulting publication had misrepresented the results.

Brown University, employer of the lead author, Martin B. Keller, launched an internal investigation, the results of which were not made public (Basken, 2012). No actions were taken against Keller, or the other 21 authors listed on the paper. Keller and at least five of the other authors continue to receive federal funding from the National Institutes of Health. *The Journal of the American Academy of Child and Adolescent Psychiatry*, which published the article, has not yet retracted it.

A recent reanalysis of Keller et al.'s 2001 study found no significant differences in efficacy between Paxil and the placebo in treating adolescents with major depression, but did find adverse emotional effects leading to increased suicidal thoughts and attempts for adolescents being treated with Paxil (Le Noury et al., 2015).

In 2015, Keller and 8 of the 22 authors of the original study wrote a letter to the blog *Retraction Watch* rebutting many points of Le Noury et al.'s 2015 reanalysis of the study; Keller claimed that data used in the reanalysis were not available during the time of the original study. He also firmly asserted that none of the paper was ghostwritten. Keller concluded that describing the original "trial as 'misreported' is pejorative and wrong," specifically from a retrospective point of view (Keller et al., 2015).

At this point, it appears that key issues related to this episode may never be resolved. In addition to the Paxil case, there have been several other cases of possible biomedical ghost writing that led to legal consequences for both medical companies and ghostwriters, indicating a heightened level of responsibility on the part of authors (see Chapter 7).

The Food and Drug Administration recently released draft guidance on publications reporting use of approved products for off-label indications: *Guidance for Industry Distributing Scientific and Medical Publications on Risk Information for Approved Prescription Drugs and Biological Products—Recommended*

*Practices*. The guidelines state that scientific journals should not publish articles "written, edited, excerpted, or published specifically for, or at the request of, a drug or device manufacturer," nor "be edited or significantly influenced by a drug or device manufacturer or any individuals having a financial relationship with the manufacturer" (FDA, 2014). In addition, articles including information on pharmaceuticals should include a statement disclosing the manufacturer's interest in the drug and any financial interest between authors and the manufacturer (FDA, 2014). Final guidance is expected, but has not yet been released.

## THE GOODWIN CASE AT THE UNIVERSITY OF WISCONSIN

***Synopsis and Rationale for Inclusion:*** *Graduate students may need support and protection from repercussions that may arise as a result of research misconduct committed by their mentor. Students stand to lose years of work if their mentor is found guilty of research misconduct, and may need to find another research group to continue their work, restart their graduate research from the beginning, or leave academia completely. With this in mind, graduate students of Elizabeth Goodwin, formerly a geneticist at the University of Wisconsin, found that data had been fabricated in one of Goodwin's proposals and reported her to the university. This case demonstrates difficult choices that may confront whistleblowers, especially those in vulnerable positions such as graduate students or postdoctoral fellows, the need for institutions to support young researchers put into difficult situations through no fault of their own, and the need for better mentoring in some laboratory and institutional environments.*

In fall 2005, graduate students working in the laboratory of University of Wisconsin geneticist Elizabeth Goodwin were confronted with evidence that their advisor had falsified data contained in a proposal to the National Institutes of Health (Couzin, 2006). Specifically, one experiment described in the proposal had not actually been performed, and figures appeared to have been manipulated. Over a period of several months, the students sought explanations from Goodwin, with which they were ultimately unsatisfied, and discussed among themselves what they should do (Allen, 2012). Recognizing that a decision to bring their concerns to university administrators would essentially shut down Goodwin's lab and have a severe negative impact on their own graduate careers, they decided that any such decision would need to be made unanimously.

Ultimately, the students decided to turn Goodwin in, which led to a university investigation finding that data in several grant applications had been falsified, a ruling confirmed by the Office of Research Integrity (ORI, 2010). Goodwin also pled guilty to making false statements on government documents, and was sentenced to 2 years' probation, fined $500, and was ordered to pay $100,000 in restitution (Winter, 2010). Several papers that Goodwin had coauthored were also investigated, but falsification was not found.

As they anticipated, the graduate students did suffer negative impacts from the case (Allen, 2012). One was able to continue work in another lab, and one was able to start a new project in a different lab at Wisconsin. One left Wisconsin to enter the PhD program at another institution, essentially starting over after 4 years. The remaining three students decided to embark on careers outside of academic research.

The case highlights several key issues. The first is the importance of whistleblowers to the system of ensuring research integrity. Although failure to replicate results, statistical analysis, and other mechanisms may be increasingly important in uncovering research misconduct, postdoctoral fellows and graduate students are responsible for reporting a significant percentage (up to half) of cases involving nonclinical research that come to ORI (Couzin, 2006). And these whistleblowers often suffer negative consequences, primarily severe damage to their careers, even when the institution takes appropriate steps to protect them from retaliation.

In addition, former students report that in the years immediately preceding Goodwin's falsified applications, problems were apparent in the lab. Several students were not making progress on their research, with no publications to show for years of work, but were advised to continue on these "dead projects" (Allen, 2012). Goodwin had also reportedly been encouraging students to overinterpret data and conceal data that conflicted with desired results (Couzin, 2006). Such ineffective mentoring and promotion of detrimental research practices create a poor environment for research integrity.

## THE HWANG STEM CELL CASE AND THE UNIVERSITY OF PITTSBURGH: COAUTHOR RESPONSIBILITIES AND INSTITUTIONAL RESPONSES

***Synopsis and Rationale for Inclusion:*** *The Hwang case raises several important research integrity issues, including data fabrication and falsification, abuse of mentorship status, whistleblower retaliation, and endangering the health of trial participants. The University of Pittsburgh's role in this case highlights the need for institutional oversight and defined standards for authorship roles. A second, more recent case at the University of Pittsburgh further demonstrates the need for oversight and institutional focus on addressing all cases of research misconduct.*

One highly publicized case that raises several important research integrity issues is that of Hwang Woo-suk, whose purportedly groundbreaking stem cell research turned out to be based on fabricated experiments (Holden, 2006). In his first article published in *Science* (in 2004), Hwang claimed to have "generated embryonic stem cells from an adult human cell," a process often referred to as therapeutic cloning, so that cells could be transplanted "without immune rejec-

tion to treat degenerative disorders" (Wade, 2006; Hwang et al., 2004, retracted). University of Pittsburgh stem cell researcher Gerald Schatten began corresponding with Dr. Hwang in late 2003, offering editorial input and support to Hwang's 2004 paper that had earlier been rejected by *Science*. Following the acceptance of the paper, Schatten and Hwang began discussing a follow-up paper in which Hwang claimed his laboratory team had "created human embryonic stem cells genetically matched to specific patients" (Sang-Hun, 2009). According to Schatten, he and Hwang drafted and edited the article together; Schatten was responsible for much of the writing and was a prominent public promoter of the findings (University of Pittsburgh, 2006). The article was published in *Science* in 2005 naming Schatten as a senior author, a role he later denied, claiming to have been no more than a coauthor.

In June 2005, immediately following the second article's published release and Hwang's announcement of a clinical trial, Young-Joon Ryu, a former researcher in Hwang's laboratory aware of the fabricated data, worried for the safety of trial participants. Ryu e-mailed Korean television network, Munhwa Broadcasting Corporation (MBC) recommending an investigation (Cyranoski, 2014b). Unfortunately, Ryu endured negative effects for his role as a whistleblower. Ryu's identity was leaked early on in the MBC investigation and he received negative backlash from Hwang's ardent supporters that led to Ryu's resignation from his position at a hospital and to a period of unemployment.

As the MBC investigation was under way, ethical concerns with Hwang's research methods were being raised. Sun Il Roh, a coauthor of the 2005 paper and fertility specialist at a hospital in Seoul, disclosed that 20 eggs he had provided to Hwang for the study had been paid for (a violation of human subjects protections), but that Hwang was unaware of this (Cyranoski and Check, 2005a). Amongst this and other signs that accepted ethical procedures were not being followed, including that a young, female graduate student in Hwang's laboratory had donated eggs to the experiment (another violation of human subjects standards), Schatten asked that his name be removed from the 2005 publication and ceased working with Hwang (Cyranoski and Check, 2005b). Four days after Roh came forward and after a year of denials, Hwang admitted that "his stem-cell research used eggs from paid donors and junior members of his team" (Cyranoski and Check, 2005a). Days later, Hwang revealed to *Science* that of the 11 photos used in the 2005 article, several were duplicates, "even though each was meant to show a different human cell colony" (Wade, 2005). Hwang claimed that this was a mistake and that it occurred only when *Science* requested higher-resolution photos, not in the original submission. Roh was interviewed in the MBC television broadcast on Hwang and revealed that "Hwang had told him 'there are no cloned embryonic stem cells'" (Cyranoski, 2005).

After its formal investigation in 2005, a Seoul National University committee determined that both of Hwang's articles were based on fabricated data (SNU, 2006). Numerous accusations ensued with Hwang admitting to "order-

ing subordinates to fabricate data," but also blaming a coauthor who "admitted to switching stem cells without Hwang's knowledge" (Cyranoski, 2014c). Preceding the SNU investigation's conclusion, Schatten and Hwang had together requested that the paper be retracted from *Science*. Based on the investigation findings, Donald Kennedy, *Science* editor-in-chief, retracted both the 2004 and 2005 papers, reporting that "seven of the 15 authors of Hwang et al., 2004 have agreed to retract their paper" and "all of the authors of Hwang et al., 2005 have agreed to retract their paper" (Kennedy, 2006). Following the retractions, Korea's National Bioethics Committee (created in response to ethical questions concerning Hwang's early research) found that Hwang had "forced junior members of his lab to donate eggs, and that he used more than 2,221 eggs in his research" (*Nature*, 2005). Hwang had only reported using approximately 400 eggs. Throughout the entire investigation, Hwang maintained that his laboratory did "create stem cells matched to individual patients," but acknowledged that mistakes were made throughout the research process. His achievement of the first cloned dog, Snuppy, was never discredited (*Nature*, 2005).

Hwang was indicted on three charges, "embezzling KRW2.8 billion [(US$2.4 million)], committing fraud by knowingly using fabricated data to apply for research funds, and violating a bioethics law that outlaws the purchase of eggs for research" (*Nature*, 2005). In 2009, Hwang was convicted on two of the three charges, violating the bioethics law and embezzling government funds. The fraud charge was dropped because the "companies involved gave the money knowing that they would not benefit from the donation" (Cyranoski, 2014a). Hwang was sentenced to a 2-year suspended prison sentence.

Today, with private funding, Hwang runs the Sooam Biotech Research Foundation that he opened in July 2006. The laboratory clones animals with the goals of "producing drugs, curing diabetes and Alzheimer's disease, providing transplantable organs, saving endangered species and relieving grief-stricken pet owners" (Cyranoski, 2014a). Since opening Sooam, Hwang has been published in peer-reviewed journals and has been successful in obtaining a Canadian patent on a cloned cell line (NT-1), which was found to be fraudulent in Hwang's 2004 *Science* article. While Hwang attempts to make a comeback, he has twice been denied approval for therapeutic cloning of human embryos by the Korean health ministry and, for now, continues to clone animals.

While a subsequent investigation by a University of Pittsburgh panel found that Gerald Schatten had not been involved with the fabrication, the incident raised questions about whether Schatten's contributions to the paper merited authorship in the first place. To what extent should coauthors, honorary or otherwise, be held responsible for the fabricated results of their collaborators? Schatten argued over the definition of the term *write*, as he did not generate the data on which the text was based, but the panel found this and disagreements over the definition of *senior author* to be dishonest attempts to relieve himself of responsibility (University of Pittsburgh, 2006). The panel found Schatten's authorship

role to be reasonable given that he wrote each draft of the paper. Schatten was also named coauthor on Hwang's 2005 Snuppy paper; however, Schatten reported to the panel that his "major contribution to the paper" was to suggest using a professional photographer to present Snuppy (University of Pittsburgh, 2006). The panel did not doubt this claim, but found it "less clear that this contribution fully justified co-authorship" (University of Pittsburgh, 2006). At his own request, Schatten was not acknowledged in Hwang's 2004 paper. Among questions of the appropriateness of authorship, also ethically problematic was Schatten's acceptance of approximately $40,000 in honoraria and research proposals to Hwang's laboratory valued at more than $200,000 for a 4-month period with implications that the grant would be continued annually (University of Pittsburgh, 2006).

The University of Pittsburgh panel's report stated that Schatten "did not exercise a sufficiently critical perspective as a scientist," but because he likely did not "intentionally falsify or fabricate experimental data, and there is no evidence that he was aware of the misconduct," he was found guilty of "research misbehavior" rather than "research misconduct" (University of Pittsburgh, 2006). "Research misbehavior" was not used or defined in the University of Pittsburgh research misconduct policy in effect at the time. The panel did not recommend any specific disciplinary action against him. Chris Pascal, director of the Office of Research Integrity supported the decision, stating "universities have a right to add refinements to categories of malfeasance" (Holden, 2006). The term *research impropriety* is contained in the University of Pittsburgh research misconduct policy adopted in 2008 (University of Pittsburgh, 2008).

## THE TRANSLATIONAL OMICS CASE AT DUKE

***Synopsis and Rationale for Inclusion:*** *The case of Duke University researchers Joseph Nevins and Anil Potti, which stretched out over several years and attracted national media attention, illustrates shortcomings and deficiencies in current approaches to research integrity on the part of researchers, research institutions, government agencies and journals (CBS News, 2012). Potti's fabricated results endangered trial participants and may have contributed to public mistrust in scientific research. Institutionally, supervisors at the laboratory level and senior administrators did not respond effectively for several years despite multiple warning signs. This case also raises questions about the responsibility of a journal to respond appropriately if numerous inquiries are made on the same original article. Several parties' unresponsiveness to questions on Potti's work may have delayed the findings of research misconduct.*

*Omics* is the study of molecules in cells, such as DNA sequences (genomics) and proteins (proteomics). Translational omics research seeks to apply this new knowledge to the creation of diagnostic tests that better detect disease and deter-

mine individualized treatment. Translational omics involves several significant challenges. Research "generates complex high-dimensional data" and resulting diagnostics are characterized by "difficulty in defining the biological rationale . . . based on multiple individual biomarkers" (IOM, 2012). In addition, diagnostic tests differ from drugs and other medical technologies regarding regulatory oversight; tests may be reviewed by the Food and Drug Administration, or be validated in a CLIA-certified laboratory (Clinical Laboratory Improvement Act).

Beginning in 2006, a series of papers appearing in major journals such as *Nature Medicine* and the *New England Journal of Medicine* purported to show that the gene activity in a patient's tumor cells could be used to determine which chemotherapy drugs would be most effective for that patient. This capability would enable significant advances in cancer treatment. Since individual reactions to these drugs are heterogeneous, the drugs that are effective for one person may not be effective for another. The lead author of the papers was cancer researcher Anil Potti, who worked at Duke University in the lab of Joseph Nevins.

Soon after the first papers were published, Keith Baggerly, Kevin Coombes, and Jing Wang, bioinformaticians at the M. D. Anderson Cancer Center of the University of Texas, began working to replicate the results. They immediately encountered difficulties using the data made publicly available with the paper, and began communicating with Potti and Nevins. Data provided by the Duke team to Baggerly, Coombes, and Wang contained numerous anomalies and obvious errors, making it impossible to replicate or verify the results. A correspondence by the M. D. Anderson researchers submitted to *Nature Medicine* in 2007 raising these issues was quickly rebutted by Potti and Nevins (Coombes et al., 2007; Potti and Nevins, 2007). However, when Baggerly, Coombes, and Wang examined additional information provided by the Duke team they found that there were still significant problems. For example, in some cases, sensitive and resistant labels for cell lines were reversed, which would lead to patients being treated with the least effective chemotherapy drug if the tests were used to direct treatment, rather than the most effective.

Over the next several years, in response to interest expressed by M. D. Anderson clinicians in utilizing the advances that continued to be reported by Potti and Nevins, Baggerly and Coombes worked with the data. In several cases where they discovered clearly incorrect results, they submitted correspondence to journals such as *Lancet Oncology*, *Journal of Clinical Oncology*, and *Nature Medicine*, but these were rejected without explanation (Baggerly, 2010, 2012).

In 2007, at the same time questions were being raised about the data underlying the Nevins-Potti research, Duke University and Duke University Medical Center investigators not associated with Nevins or Potti launched three clinical trials based on the results, and an additional trial was launched at Moffitt Cancer Center (IOM, 2012). Duke also applied for patents, and several companies were working to commercialize the research, including one in which Potti served as a director and secretary (Reich, 2010b; Tracer, 2010). Learning about the trials in

June 2009, Baggerly and Coombes prepared a critical analysis of the Duke work, which was published in the *Annals of Applied Statistics* after it had been rejected by a biomedical journal (Baggerly and Coombes, 2009).

In January 2015, the *Cancer Letter*, a specialist newsletter, reported that Bradford Perez, a third-year medical student who was working with Potti in the Nevins lab, became very concerned about the methodology and reliability of the research (Goldberg, 2015). He shared these concerns in a detailed memo with Potti, Nevins, and several Duke administrators in the spring of 2008 (Goldberg, 2015). In addition to providing specifics about a number of concerning factors, he asked that his name be removed from four papers based on the work he had contributed to, including a paper submitted to the *Journal of Clinical Oncology*, and left the Nevins-Potti laboratory (Perez, 2008). Rather than catalyzing any independent assessment of the serious concerns raised by Perez about the quality of the research, Duke administrators referred him back to Nevins with no apparent follow-up by any institutional official. Nevins and Potti committed to revalidate all of their work, but it appears that this did not happen. Perez left the Nevins lab knowing he would repeat a year of his medical education, in his words, "to gain a more meaningful research experience" (Perez, 2008).

As noted in a 2012 Institute of Medicine (IOM) report discussed further below, Duke "did not institute extra oversight or launch formal investigations of the three trials during the first 3 years after the original publications triggered widely known controversy about the scientific claims and after concerns started to develop about the possible premature early initiation of clinical trials" (IOM, 2012). Not only did Duke's administration fail to act decisively on Perez's suspicions, but an administrator who counseled Perez on the matter did not even inform the IOM committee that Perez had come forward years earlier (Goldberg, 2015; IOM, 2012). In response to the 2015 revelations by the *Cancer Letter,* Duke Medicine officials did not answer specific questions, but did state that "there are many aspects of this situation that would have been handled differently had there been more complete information at the time decisions were made" (Goldberg, 2015).

National Cancer Institute (NCI) researcher Lisa McShane had also been unsuccessful in attempts to replicate the work (*Economist*, 2013). In the fall of 2009, NCI expressed concern about the clinical trials at Duke as well as the parallel trial at Moffitt. The trials were suspended, and Duke's Institutional Review Board formed an external review panel to evaluate the concerns. The Duke trials were restarted in early 2010 after the review panel concluded that the approaches used in the trials were "viable and likely to succeed" (IOM, 2012).

During the first half of 2010, NCI continued to raise questions about the research. Through a Freedom of Information Act request *submitted by* the *Cancer Letter,* it was revealed that the external review panel was not provided with several critical pieces of information, including a detailed description of the statistical methods used in the original research, and a new critique from Baggerly and

Coombes based on analysis of updated data posted by Potti and Nevins (Baggerly, 2010; Duke University, 2009). About that material, the 2012 IOM report notes that it "was never forwarded to the external statistical reviewers because of the university leadership's concerns that it might 'bias' the committee's review" (IOM, 2012).

Several developments in July 2010 brought matters to a head. It was reported that Potti's claim on his resume that he had been a Rhodes Scholar was exaggerated, and this was confirmed by the University of Oxford (Goldberg, 2010; Singer, 2010). Also, several dozen prominent biostatisticians wrote to NCI director Harold Varmus to request that the clinical trials based on the Duke research be suspended until the science could be publicly clarified (Barón et.al., 2010; Singer, 2010). Duke suspended the trials and suspended Anil Potti's employment in response. The trials were ultimately terminated and Potti left Duke. Starting in the fall of 2010, a number of the papers reporting the Duke results have been retracted.

Over the time since the trials were suspended, there have been several significant developments. NCI asked the Institute of Medicine to develop principles for evaluating omics-based tests, and IOM released its report in 2012 (IOM, 2012). Drawing on lessons from the Duke case and informed by the development of other omics-based tests, the IOM report lays out a recommended development and evaluation process for these tests, and makes specific implementation recommendations to researchers, institutions, agencies, and journals (IOM, 2012).

Duke University has also taken steps to respond (Califf, 2012). Its Translational Medicine Quality Framework emphasizes new science and management approaches to ensure data provenance and integrity, the incorporation of adequate quantitative expertise, explicit management accountability in the institution beyond the individual lab for research affecting patient care, and enhanced conflict-of-interest reviews.

In 2015, ORI concluded that Potti had "engaged in research misconduct by including false research data," citing specific examples of Potti's data that had been reversed, switched, or changed in a number of (now retracted) articles and other submissions (ORI, 2015). While Potti did not "admit nor deny ORI's findings of research misconduct," he has expressed that he has no intention of applying for PHS (Public Health Service)–funded research, but agreed that if he is engaged with any PHS-funded research in the future, his research will be supervised for 5 years (ORI, 2015).

In this case, just about all the scientific checks and balances intended to uncover incorrect or fabricated research and protect human subjects failed over the course of several years. A summary of these failings illustrates some of the U.S. research enterprise's key vulnerabilities regarding integrity. Effective steps on the part of Duke to address the problems with Potti's work and investigate possible misconduct were delayed for years, and were finally triggered only by the disclosure of Potti's resume falsification. Those pointing out these problems

were appropriately cautious about making formal allegations of misconduct, since there was a possibility that the problems were due to error or extreme sloppiness rather than falsification. Another contributing factor was the willingness of Joseph Nevins, a highly prestigious researcher, to vouch for the work and advocate for Potti with university administrators and others.

### Individual Researchers

Anil Potti's misbehavior is at the center of the case. Prior to ORI's conclusion of research misconduct, Joseph Nevins and Robert Califf had both said that it is highly likely that Potti intentionally fabricated or falsified data (CBS News, 2012). In addition, Baggerly, Coombes, and Wang had documented many instances of sloppy or careless data analysis, and Perez documented use of unreliable predictors and omission of data not showing desired results. The negative impact of such sloppy and careless practices on the ability to replicate results and ultimately on patient care might be similar to the impact of fabrication or falsification.

In addition to problems with data and analysis, the IOM committee described a number of poor practices related to the clinical trials for the tests, including trials being undertaken simultaneously with preliminary studies (IOM, 2012).

Potti's collaborators also share responsibility. For example, despite being principal investigator of the lab where the research was undertaken, as well as Potti's mentor and coauthor, Joseph Nevins did not thoroughly check the original data files until after it was revealed that Potti had exaggerated his credentials in July 2010, more than 3 years after the data issues were originally raised (CBS News, 2012). Moreover, we now know from a deposition cited in court documents that Nevins "pleaded with Perez not to send a letter about his concerns to the Howard Hughes Medical Institute, which was supporting him, because it would trigger an investigation at Duke" (Kaiser, 2015). Indeed, Duke administrators testified to the IOM that none of Potti's coauthors (a total of 162 for 40 papers) raised any questions or concerns about the papers or tests until they were contacted by Duke at the start of the process of determining which papers should be retracted (IOM, 2012). Bradford Perez, the medical student described above, did raise concerns and removed his name from the papers that he contributed to, so his documented concerns were apparently not considered when that statement was made. Nevins remained on faculty as a department chair until his retirement in 2013, the year after the IOM report was released.

### Institutional Policies and Procedures

In addition to the failures of individual researchers, lessons can be drawn from the responses by Duke as an institution during the controversy. Institutional shortcomings in policies and procedures, structure, systems, and oversight

contributed to delays in recognizing that the science underlying the Nevins-Potti research was unsound. First, Duke's Institute for Genomic Science and Policy and its component Center for Applied Genomics and Technology, where Nevins and Potti worked, instituted its own system for undertaking clinical trials, separate from the extensive existing infrastructure of the Duke Cancer Center (IOM, 2012). This parallel pathway lacked the normal checks and balances as well as clear lines of authority and oversight.

In addition, systems for managing conflicts of interest at the individual and institutional levels were inadequate (IOM, 2012). For example, the IOM committee found evidence that researchers involved with undertaking the clinical trials had unreported financial or professional conflicts of interest. Some investigators held patents on one or more of the tests, or had links with one of the companies founded to market the tests. The institution itself, through its licensing relationships, had a financial interest in the success of the tests, as well as a reputational interest in having generated such an important new technology. It is of note that the institution had created a set of video and print materials featuring the research (CBS News, 2012; Singer, 2010).

As noted in the 2012 IOM report, as a "responsible party" for assuring the integrity of the science conducted under their auspices, universities have particularly important responsibilities. These include responsibility for the hiring and promotion of the faculty members conducting research, the establishment and maintenance of oversight structures, and responsibilities for properly responding to and resolving questions about the validity of research or allegations of misconduct when they arise. It also includes the responsibility for ensuring the existence of an organizational culture and climate that sets expectations for research integrity that "are transmitted by the institution and modeled by its leadership. Institutional culture starts with the dean, senior leaders, and members of their team stating how research is to be conducted, with integrity and transparency, and with clarity that shortcuts will not be tolerated and that dishonesty is the basis for dismissal" (IOM, 2012).

The evidence now available, some that has come to light only after Freedom of Information Act requests and court depositions, suggests that Duke University and its leadership failed in virtually all of these responsibilities: for undertaking clinical studies outside the established review structures; for the failure to pursue internal investigation of serious, documented concerns until forced by outside forces to do so; for withholding from an external committee the full Baggerly/Coombes critique; for referring responsibility for rechecking Potti's work back to the laboratory of his (explicitly conflicted) principal investigator, Joseph Nevins; for failing to employ the full set of institutional checks and balances that were in place; and for either incomplete or factually unsupportable statements made to the IOM Committee charged with examining the issue. The breadth and depth of these institutional failings are disappointing. Occurring in an institution of Duke's stature and resources, they raise troubling questions about the ability of research

institutions, without more support and reinforcement, to manage complex cases when directed against prominent institutional researchers.

Duke suspended the trials and launched an investigation in the fall of 2009 in response to NCI concerns. However, this investigation had several serious flaws. Although the trials were resumed based on the report of the two external statistical experts, as noted above, these experts were not provided with several critical pieces of information. The IOM report also raises the possibility that Nevins was improperly in direct contact with the reviewers during the inquiry (IOM, 2012). As for the clinical trials that were undertaken based on the fabricated work, 117 patients were ultimately enrolled. Duke later faced a lawsuit brought by the families of eight of these patients, which was settled in May 2015. The terms of the settlement were not disclosed (Ramkumar, 2015).

In its Translational Medicine Quality Framework activity, Duke also identified an environment that might discourage postdocs or grad students from raising concerns with research within the lab or taking their concerns to others at the university as a possible problem. The university reported that it has established an ombudsman's office and taken other steps to address this.

Taken together, these institutional failings raise the question of whether, in addition to strengthening policies and procedures to the extent possible, research institutions should explore new mechanisms for bringing in outside perspectives in cases where it might be difficult for an institution to objectively address allegations of misconduct or other challenges to the soundness of science. In 2016, four members of the IOM committee published a piece critical of how Duke handled the case as an institution (DeMets et al., 2016).

### Journal Policies and Practices

Although *Nature Medicine* and the *Journal of Clinical Oncology* did publish letters from Baggerly, Coombes, and Wang questioning the validity of data, along with responses from Potti, they rejected further questioning of the Duke results. This is likely the result of the common journal practice of not publishing additional comments on an article that appear to repeat concerns already raised in a previously published comment, so as to avoid involving the journal in an ongoing dispute. Further, other journals that had published other articles reporting the Nevins-Potti work were not responsive to questions raised by Baggerly and Coombes. This stance contributed to delays in recognizing the nature and extent of the problems with the papers. The translational omics case raises issues of how scholarly publishers, institutions, and the broader community should respond when the work underlying numerous papers in a variety of journals is questioned.

### Sponsor and Regulator Policies and Practices

The IOM report identifies some ambiguities in Food and Drug Administration requirements for launching clinical trials on diagnostics as possibly contributing to the clinical trials being launched prematurely and to delays in finally shutting them down (IOM, 2012). The IOM report also points out that NCI felt constrained in communicating what it knew and the extent of its concerns with Duke and others early in the case, particularly before officials were aware that the agency was supporting aspects of the clinical trials (IOM, 2012). More direct and complete communication would be helpful in future cases.

## THE RIKEN-STAP CASE

***Synopsis and Rationale for Inclusion:*** *The RIKEN-STAP case illustrates issues that may arise related to authorship roles, mentoring, and data falsification. The extent to which coauthors should be held responsible for the data and findings of papers on which they are listed is a recurring question in many research misconduct cases.*

Yoshiki Sasai, a stem cell biologist of Japan's RIKEN research institute, committed suicide in August 2014 after the lead author on papers that he coauthored, Haruko Obokata, was found guilty of research misconduct (RIKEN, 2014). Obokata claimed to have found that a process that reprogrammed somatic cells into pluripotent cells by exposing the cells to stress; the authors termed the process "stimulus-triggered acquisition of pluripotency (STAP)" (Obokata et al., 2014a, retracted). Obokata collaborated with Charles Vacanti's laboratory at Brigham and Women's Hospital, where the idea of STAP had supposedly originated (Knoepfler, 2015). Vacanti, professor of anesthesiology at Harvard Medical School and former chairman of the Department of Anesthesia at the Brigham and Women's Hospital, was a corresponding author on one of the papers, a coauthor on the other, and Obokata's mentor while she worked as a postdoctoral research fellow at Brigham and Women's Hospital.

Shortly after Obokata's findings were published in *Nature*, outside researchers were unable to replicate the study or achieve similar results, prompting an internal RIKEN investigation. The investigation committee concluded that she had fabricated data in at least one of the papers (RIKEN, 2014). The committee found problems with the data underlying the other papers, but was not able to conclude that fabrication or falsification had occurred because they did not have access to the original data (RIKEN, 2014). The committee found that Sasai had no involvement with the data fabrication, but bore a "heavy responsibility" for the incident because he did not insist that experiments be repeated even after problems with the data became obvious (RIKEN, 2014)

Both Sasai and Obokata made public apologies, but maintained that STAP works. Already disgraced, the Japanese media soon began to make "unsubstanti-

ated claims about [Sasai's] motivations" and personal life, as well as shame him for a lack of oversight responsibility, all of which, Sasai wrote in a suicide note, drove him to take his own life (Cyranoski, 2014c). Vacanti also maintained "absolute confidence" in the phenomenon and released follow-up protocols to the retracted *Nature* papers to assist in the reproducibility of STAP cells (Vacanti and Kojima, 2014). Following RIKEN's investigation and the retraction of the *Nature* papers, Vacanti stepped down as chairman of the Department of Anesthesia at the Brigham and Women's Hospital and took a 1-year sabbatical from his professorship at Harvard Medical School. He did not reference the STAP case in his letter of resignation from Brigham and Women's Hospital.

## REFERENCES FOR APPENDIX D


Allen, M. 2012. The Goodwin Lab. Presentation to the committee. August 14.

Baggerly, K. A. 2010. *The Importance of Reproducible Research in High-Throughput Biology: Case Studies in Forensic Bioinformatics*. Video Lecture. Available at: http://videolectures.net/cancerbioinformatics2010_baggerly_irrh/. Accessed November 20, 2016.

Baggerly, K. A. 2012. Learning from the Duke Case and the IOM Translational Omics Report: Context. Presentation to the committee. July 9.

Baggerly, K. A., and K. R. Coombes. 2009. Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology. *Annals of Applied Statistics* 3(4): 1309-1334.

Baron, A. E., K. Bandeen-Roche, D. A. Berry, J. Bryan, V. J. Carey, K. Chaloner, M. Delorenzi, B. Efron, R. C. Elston, D. Ghosh, J. D. Goldberg, S. Goodman, F. E. Harrell, S. Galloway Hilsenbeck, W. Huber, R. A. Irizarry, C. Kendziorski, M. R. Kosorok, T. A. Louis, J. S. Marron, M. Newton, M. Ochs, J. Quackenbush, G. L. Rosner, I. Ruczinski, S. Skates, T. P. Speed, J. D. Storey, Z. Szallasi, R. Tibshirani, and S. Zeger. 2010. Letter to Harold Varmus: Concerns About Prediction Models Used in Duke Clinical Trials. Bethesda, MD, July 19. http://www.cancerletter.com/categories/documents (accessed March 5, 2013).

Basken, P. 2012. Academic Researchers Escape Scrutiny in Glaxo Fraud Settlement. *Chronicle of Higher Education*. August 6. Available at http://www.chronicle.com/article/Academic-Researchers-Escape/133325/.

*BMJ* (*British Medical Journal*). 2011. Wakefield's article linking MMR vaccine and autism was fraudulent. January 6. Available at: https://doi.org/10.1136/bmj.c7452; Accessed November 12, 2016.

Bosch, X., and J. Ross. 2012. Ghostwriting: Research misconduct, plagiarism, or fool's gold? *American Journal of Medicine* 125(4): 324-326.

Boseley, A. 2010. *Lancet* retracts "utterly false" MMR paper. *The Guardian*. February 2. Available at: http://www.theguardian.com/society/2010/feb/02/lancet-retracts-mmr-paper.

Califf, R. M. 2012. Translational Medicine Quality Framework: Challenges and Opportunities. Presentation to the committee. July 9.

CBS News. 2012. Deception at Duke. 60 Minutes. February 12.

Coombes, K. R., J. Wang, and K. A. Baggerly. 2007. Microarrays: Retracing steps. *Nature Medicine* 13: 1276-1277.

COPE (Committee on Publication Ethics). 2011. Code of Conduct and Best Practice Guidelines for Journal Editors. Available at: http://publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf. Accessed July 23, 2013.

Couzin, J. 2006. Truth and consequences. *Science* 313: 1222-1226.

Cyranoski, D. 2005. Stem-cell pioneer accused of faking data. *Nature*. December 15. doi:10.1038/news051212-14.

Cyranoski, D. 2014a. Cloning comeback. *Nature* 505(7484). January 23. doi:10.1038/505468a.

Cyranoski. D. 2014b. Whistle-blower breaks his silence. *Nature* 505(7485). January 30. doi:10.1038/505593a.

Cyranoski, D. 2014c. Stem-cell pioneer blamed media 'bashing' in suicide note. *Nature*. August 13. Available at: http://www.nature.com/news/stem-cell-pioneer-blamed-media-bashing-in-suicide-note-1.15715.

Cyranoski, D. and E. Check. 2005a. Clone star admits lies over eggs. *Nature* 438: 536-537. December 1. doi:10.1038/438536a.

Cyranoski, D., and E. Check. 2005b. Stem-cell brothers divide. *Nature* 438: 262-263. November 16. doi:10.1038/438262a.

Deer, B. 2011a. Exposed: Andrew Wakefield and the MMR-autism fraud. http://briandeer.com/mmr/lancet-summary.htm.

Deer, B. 2011b. How the case against the MMR vaccine was fixed. *BMJ* 342: 77-82.

Deer, B. 2014. Disgraced ex-doctor fails in his fourth attempt to gag media. http://briandeer.com/solved/slapp-introduction.htm.

DeMets, D. L., T. R. Fleming, G. Geller, and D. F. Ransohoff. 2016. Institutional responsibility and the flawed genomic biomarkers at Duke University: A missed opportunity for transparency and accountability. *Science and Engineering Ethics* November 23. doi: 10.1007/s11948-016-9844-4.

Duke University. 2009. Review of Genomic Predictors for Clinical Trials from Nevins, Potti, and Barry.

*Economist, The*. 2013. Trouble at the lab. October 19. Available at: https://www.economist.com/news/briefing/21588057-scientists-think-science-self-correcting-alarming-degree-it-not-trouble.

Eggertson, L. 2010. *Lancet* retracts 12-year-old article linking autism to MMR vaccines. *Canadian Medical Association Journal* 182(4): 188-200.

FDA (Food and Drug Administration). 2014. (Draft) Guidance for Industry: Distributing Scientific and Medical Publications on Risk Information for Approved Prescription Drugs and Biological Products—Recommended Practices. Available at: http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm400104.pdf.

Goldberg, P. 2015. Duke officials silenced med student who reported trouble in Anil Potti's lab. *Cancer Letter* 41: 1 (January 9).

Goldberg, P. 2010. Prominent Duke scientist claimed prizes he didn't win, including Rhodes Scholarship. *The Cancer Letter* 36(27): 1-7.

Holden, C. 2006. Schatten: Pitt panel finds "misbehavior" but not misconduct. *Science* 311: 928.

Hooker, B. S. 2014. Measles-mumps-rubella vaccination timing and autism among young African American boys: A reanalysis of CDC data. *Translational Neurodegeneration* 3: 16. RETRACTED.

Hwang, W. S., J. R. Young, J. H. Park, E. S. Park, E. G. Lee, J. M. Koo, H. Y. Jeon, B. C. Lee, S. K. Kang, S. J. Kim, C. Ahn, J. H. Hwang, K. Y. Park, J. B. Cibelli, and S. Y. Moon. 2004. Evidence of a pluripotent human embryonic stem cell line derived from a cloned blastocyst. *Science* 303(5664): 1669-1674. RETRACTED.

ICMJE (International Committee of Medical Journal Editors). 2015. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. Updated December 2015. Available at: http://www.icmje.org/recommendations/. Accessed November 14, 2016.

IOM (Institute of Medicine). 2012. *Evolution of Translational Omics: Lessons Learned and the Path Forward*. Washington, DC: The National Academies Press.

Kaiser, J. 2015. Duke University officials rebuffed medical student's allegations of research problems. *Science*. January 12. doi: 10.1126/science.aaa6330.

Keller, M. B., B. Birmacher, G. N. Clarke, G. J. Emslie, H. Koplewicz, S. Kutcher, N. Ryan, W. H. Sack, and M. Strober. 2015. Letter to *Retraction Watch*. Available at: http://retractionwatch.com/wp-content/uploads/2015/12/Response-to-BMJ-Article-9-1515.pdf.

Kennedy, D. 2006. Editorial Retraction: Retraction of Hwang et al., *Science* 308(5729): 1777-1783. *Science* 311(5759): 335.

Knoepfler, P. 2015. New *Nature* papers debunk STAP cells. Knoepfler Lab Stem Cell Blog. September 23. Available at: http://www.ipscell.com/tag/charles-vacanti/.

*Lancet* Editors. 2010. Retraction—Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet* 375(9713): 445. doi: http://dx.doi.org/10.1016/S0140-6736(10)60175-4.

Le Noury, J., J. Nardo, D. Healy, J. Jureidini, M. Raven, C. Tufanaru, and E. Abi-Jaoude. 2015. Restoring study 329: Efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ* 351: h4320. Available at: https://doi.org/10.1136/bmj.h4320.

Lindell, M. 2014. Court: Andrew Wakefield, autism researcher, cannot sue in Texas. *Statesman News*. September 19.

Murch, S. H., A. Anthony, D. H. Casson, M. Malik, M. Berelowitz, A. P. Dhillon, M. A. Thomson, A. Valentine, S. E. Davies, and J. A. Walker-Smith. 2004. Retraction of an interpretation. *The Lancet* 363(9411): 750.

ORI (Department of Health and Human Services, Office of Research Integrity). 2010. Case Summary: Goodwin, Elizabeth. August 24.

ORI. 2014. *Office of Research Integrity Newsletter* 22(4).

ORI. 2015. ORI Case Summary: Potti, Anil. Available at: https://ori.hhs.gov/content/case-summary-potti-anil; Accessed November 22, 2016.

Perez, B. 2008. Letter to Howard Hughes Medical Institute. April 22. Available at: https://issuu.com/thecancerletter/docs/duke_letters_to_hhmi; Accessed November 22, 2016.

Potti, A., and J. Nevins. 2007. Reply to "Microarrays: retracing steps." *Nature Medicine* 13: 1277-1278.

Ramkumar, A. 2015. Duke lawsuit involving cancer patients linked to Anil Potti settled. *Duke Chronicle*. May 3.

Reich, E. S. 2010a. Self-plagiarism case prompts calls for agencies to tighten rules. *Nature* 768: 745.

Reich, E. S. 2010b. Troubled geneticist rebuffed by US patent office. *Nature News*. September 7. Available at: http://www.nature.com/news/2010/100907/full/news.2010.450.html. Accessed January 4, 2017.

Reich, E. S. 2011. Fresh dispute about MMR "fraud." *Nature* 479: 157-158.

RIKEN. 2014. Report on STAP Cell Research Paper Investigation. Available at: http://www3.riken.jp/stap/e/c13document52.pdf; Accessed November 8, 2016.

Sang-Hun, C. 2009. Disgraced cloning expert convicted in South Korea. *New York Times*. October 26.

Singer, N. 2010. Duke scientist suspended over Rhodes Scholar claims. *New York Times*. July 20.

SNU (Seoul National University). 2006. Text of the Report on Dr. Hwang Woo Suk. *New York Times*. January 6.

Stern, S., and T. Lemmens. 2011. Legal remedies for medical ghostwriting: Imposing fraud liability on guest authors of ghostwritten articles. *PLOS Medicine* 8(8): e1001070.

Taylor, B., E. Miller, C. P. Farrington, M. C. Petropoulos, I.Favot-Mayaud, J. Li, and P. A. Waight. 1999. Autism and measles, mumps, and rubella vaccine: No epidemiologic evidence for a causal association. *The Lancet* 353: 2026-2029.

Thomas, K., and M. Schmidt. 2012. Glaxo agrees to pay $3 billion in fraud settlement. *New York Times*. July 2.

Tracer, Z. 2010. Health care companies cut ties with Potti. *Duke Chronicle*. September 14. Available at: http://www.dukechronicle.com/article/2010/09/health-care-companies-cut-ties-potti; Accessed January 3, 2017.

*Translational Neurodegeneration* Editor and Publisher. 2014. Retraction Note: Measles-mumps-rubella vaccination timing and autism among young African American boys: a re-analysis of CDC data. *Translational Neurodegeneration*. 3: 22. October 3. Available at: https://translationalneurodegeneration.biomedcentral.com/articles/10.1186/2047-9158-3-22.

Triggle, N. 2010. MMR doctor struck from register. *BBC News*. May 24. Available at: http://news.bbc.co.uk/2/hi/health/8695267.stm; Accessed November 12, 2016.

UCL (University College London). 2012. MMR and the development of a research governance framework. *UCL News*. September 13. Available at: http://www.ucl.ac.uk/news/news-articles/1209/13092012-Governance.

UK GMC (United Kingdom General Medical Council). 2010. Fitness to Practice Panel Hearing. 28 January.

University of Pittsburgh. 2006. Summary Investigative Report on Allegations of Possible Scientific Misconduct on the Part of Gerald P. Schatten, Ph.D. February 8.

University of Pittsburgh. 2008. Research Integrity Policy. Available at: https://www.cfo.pitt.edu/policies/policy/11/11-01-01.html.

Vacanti, C. A., and K. Kojima. 2014. Revised STAP Cell Protocol. Brigham and Women's Hospital. Available at: http://research.bwhanesthesia.org/research-groups/cterm/stap-cell-protocol.

Wade, N. 2005. Journal defends stem cell article despite photo slip. *New York Times*. December 7.

Wade, N. 2006. Journal faulted in publishing Korean's claims. *New York Times*. November 29.

Wakefield, A. J., S. H. Murch, A. Anthony, J. Linnell, D. M. Casson, M. Malik, M. Berelowitz, A. P. Dhillon, M. A. Thomson, P. Harvey, A. Valentine, S. E. Davies, and J. A. Walker-Smith (1998). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet* 351(9103): 637-641. RETRACTED.

Winter, S. 2010. Former Wisconsin researcher sentenced for misconduct. *BioTechniques*. September 17. Available at: www.biotechniques.com/news/Former-Wisconsin-researcher-sentenced-for-misconduct/biotechniques-302891.html; Accessed December 5, 2016.

Ziv, S. 2015. Andrew Wakefield, father of the anti-vaccine movement, responds to the current measles outbreak for the first time. *Newsweek*. February 10.

# Appendix E

# Bibliography

AAAS (American Association for the Advancement of Science). 2012. *Connecting Science, Engineering, Ethics and Human Rights: Beyond Human Subjects Research*. Workshop report. Washington, DC: AAAS. Available at: https://www.aaas.org/sites/default/files/migrate/uploads/Workshop-Intersection-Science-Ethics-Human-Rights.pdf.

AAMC (Association of American Medical Colleges). 2006. *Compact Between Postdoctoral Appointees and Their Mentors*. Available at: https://www.aamc.org/download/49852/data/postdoccompact.pdf. Accessed July 15, 2013.

AAMC. 2008. *Compact Between Biomedical Graduate Students and Their Research Advisors*. Available at: https://www.aamc.org/download/49868/data/gradcompact.pdf. Accessed July 15, 2013.

Aarts, A. A. C. J. Anderson, J. Anderson, M. A. L. M. van Assen, P. R. Attridge, A. S. Attwood, J. Axt, et al. 2015. Reproducibility Project: Psychology. Open Science Framework. December 8. Retrieved from osf.io/ezcuj.

Abbott, B. P., et al. 2016. Observation of gravitational waves from a binary black hole merger. *Physical Review Letters* 116; doi: 10.1103/PhysRevLett.116.061102.

Academy of Medical Sciences, Biotechnology and Biological Sciences Research Council, Medical Research Council, and Wellcome Trust. 2015. *Reproducibility and Reliability of Biomedical Research: Improving Research Practice*. London: Academy of Medical Sciences. Available at: https://acmedsci.ac.uk/viewFile/56314e40aac61.pdf.

Adams, D., and K. D. Pimple. 2005. Research misconduct and crime lessons from criminal science on preventing misconduct and promoting integrity. *Accountability in Research* 12(3): 225-240.

Agnew, R. 1992. Foundation for a general strain theory of crime and delinquency. *Criminology* 30: 47-87.

Agnew, R. 2006. General strain theory: Current status and directions for further research. Pp. 101-123 in *Taking Stock: The Status of Criminological Theory*. Edison, NJ: Transaction.

Akers, R. L. 1985. *Deviant Behavior: A Social Learning Approach*. Belmont, CA: Wadsworth.

Akers, R. L., and J. K. Cochran. 1985. Adolescent marijuana use: A test of three theories of deviant behavior. *Deviant Behavior* 6: 323-346.

Alberts, B., J. Halpern, P.H. Raven, D.D. Brown, J.E. Dowling, I.M. Singer, and M.F. Dresselhaus for the Council of the National Academy of Sciences. 1996. Letter to William Raub (March 15). Released by the National Academy of Sciences.

Alberts, B. 2012. The end of "small science"? *Science* 337(6102): 1583.

Alberts, B. 2013. Impact factor distortions. *Science* 340(6134): 787.

Alberts, B. M., M. W. Kirschner, S. Tilghman, and H. Varmus. 2014. Rescuing US biomedical research from its systemic flaws. *Proceedings of the National Academy of Sciences* 111(16): 5773-5777.

Alberts, B., R. J. Cicerone, S. E. Fienberg, A. Kamb, M. McNutt, R. M. Nerem, R. Schekman, R. Shiffrin, V. Stodden, S. Suresh, M. T. Zuber, B. K. Pope, and K. H. Jamieson. 2015. Self-correction in science at work. *Science* 348(6242): 1420-1422.

Allen, J. 2008. Can of worms. *On Wisconsin* 109(1): 28-33, 63.

Altman, E., and P. Hernon (eds). 1997. *Research Misconduct: Issues, Implications, and Strategies*. Greenwich, CT: Ablex.

Altman, L., and W. J. Broad. 2005. Global trend: More science, more fraud. *New York Times*, December 20, pp. F1, F6.

America COMPETES Reauthorization Act of 2015, H.R. 1806, 114th Congress, 1st Session. Available at: www.congress.gov/bill/114th-congress/house-bill/1806/text. Accessed December 17, 2015.

American Innovation and Competitiveness Act of 2016, H.R. 3084, 114th Congress. Available at: https://www.congress.gov/bill/114th-congress/senate-bill/3084.

American Law Institute. 1985. *Model Penal Code: Official Draft and Explanatory Notes*. Philadelphia, PA: American Law Institute.

Anderson, G., H. Sprott, and B. R. Olsen. 2013. Opinion: Publish negative results. *The Scientist*. January 15. Available at: http://www.the-scientist.com/?articles.view/articleNo/33968/title/Opinion--Publish-Negative-Results/.

Anderson, M. S. 2007. Collective openness and other recommendations for the promotion of research integrity. *Science and Engineering Ethics* 13(4): 387-394.

Anderson, M. S., A. S. Horn, K. R. Risbey, E. A. Ronning, R. De Vries, and B. C. Martinson. 2007a. What do mentoring and training in the responsible conduct of research have to do with scientists' misbehavior? Findings from a National Survey of NIH-funded scientists. *Academic Medicine* 82(9): 853-860.

Anderson, M., E. Ronning, R. De Vries, and B. Martinson. 2007b. The perverse effects of competition on scientists' work and relationships. *Science and Engineering Ethics* 13(4): 437-461.

Anderson, M. S., E. A. Ronning, R. De Vries, and B. C. Martinson. 2010. Extending the Mertonian norms: Scientists' subscription to norms of research. *Journal of Higher Education* 81: 366-393.

Antes, A. L., R. P. Brown, S. T. Murphy, E. P. Waples, M. D. Mumford, S. Connelly, and L. D. Davenport. 2007. Personality and ethical decision-making in research: The role of perceptions of self and others. *Journal of Empirical Research on Human Research Ethics* 2(4): 15-34.

Antes, A., S. T. Murphy, E. P. Waples, M. D. Mumford, R. P. Brown, S. Connelly, and L. D. Devenport. 2009. A meta-analysis of ethics instruction effectiveness in the sciences. *Ethics & Behavior* 19(5): 379-402.

Antes, A. L., X. Wang, M. D. Mumford, R. P. Brown, S. Connelly, and L. D. Devenport. 2010. Evaluating the effects that existing instruction on responsible conduct of research has on ethical decision making. *Academic Medicine* 85(3): 19-26.

APS (American Physical Society). 2004. *Ethics Case Studies: A Set of Ethics Case Studies from Physics Research, with an Activity and Discussion Guide for Courses and Seminars in Science Ethics*. Available at: http://www.aps.org/programs/education/ethics/upload/Ethics-Case-Studies-Student-Edition.pdf.

Ariely, D. 2012. *The (Honest) Truth About Dishonesty: How We Lie to Everyone—Especially Ourselves*. New York: Harper.

ASCB (American Society for Cell Biology). 2012. San Francisco Declaration on Research Assessment (DORA). Available at: http://am.ascb.org/dora/. Accessed December 15, 2014.

AUTM (Association of University Technology Managers). 2013. *AUTM U.S. Licensing Activity Survey Highlights*. Washington, DC: AUTM.

AUTM. 2014. *AUTM Annual Report*. Washington, DC: AUTM.

Azoulay, P., J. S. Graff Zivin, and G. Manso. 2011. Incentives and creativity: Evidence from the academic life source. *RAND Journal of Economics* 42(3): 527-554.

Bandura, A. 1979. The social learning perspective: Mechanisms of aggression. In *The Psychology of Crime and Criminal Justice*. New York: Holt, Rinehart & Winston.

Bannerman, D. 2014. VHA Handbook 1058.02 Research Misconduct [PowerPoint slides]. Available at: http://www.va.gov/ORO/Docs/Presentations/vhahb1058_02_pres.pdf.

Baumeister, R. F., M. Muraven, and D. M. Tice. 2000. Ego depletion: A resource model of volition, self-regulation, and controlled response. *Social Cognition* 18: 130-150.

Baumeister, J., and R. Tierney, 2011.*Willpower: Rediscovering the Greatest Human Strength*. New York: Penguin Books.

Beall, J. 2016. Scholarly Open Access: Critical analysis of scholarly open-access publishing. Blog. Available at: https://scholarlyoa.com/.

Beasley, M. R., S. Datta, H. Kogelnik, H. Kroemer, and D. Monroe. 2002. *Report of the Investigation Committee on the Possibility of Scientific Misconduct in the Work of Hendrik Schön and Coauthors*. Lucent Technologies. Available at: http://doc.sciencenet.cn/upload/file/2011922112725742.pdf.

Bebeau, M. J. 2002. The defining issues test and the four component model: Contributions to professional education. *Journal of Moral Education* 31(3): 271-295.

Bebeau, M. J., K. D. Pimple, and K. M. T. Muskavitch. 1995. *Moral Reasoning in Scientific Research: Cases for Teaching and Assessment*. I. University, Editor. Bloomington: University of Indiana. https://eric.edu.gov/?id=ED39483.

Begley, C. G., and L. M. Ellis. 2012. Raise standards for preclinical cancer research. *Nature* 483: 531-533.

Begley, C. G., and J. P. Ioannidis. 2015. Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research* 116(1):116-126.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* 57(1): 289-300.

Ben-Yehuda, N. 1986. Deviance in science: Towards the criminology of science. *British Journal of Criminology* 26(1), 1-27.

Benos, D. J., E. Bashari, J. M. Chaves, A. Gaggar, N. Kapoor, M. LaFrance, R. Mans, D. Mayhew, S. McGowan, A. Polter, Y. Qadri, S. Sarfare, K. Schultz, R. Splittgerber, J. Stephenson, C. Tower, R. G. Walton, and A. Zotov. 2007. The ups and downs of peer review. *Advances in Physiology Education* 31: 145-152.

Berger, E., and T. Ackerman. 2012. Cancer agency agrees to freeze new grants. *Houston Chronicle*. December 19.

Bienenstock, A. 2012. The Federal Research Misconduct Definition. Presentation to the Committee, Palo Alto, CA, July 9.

Bierer, B., M. Barnes, IRB/RIO/IO Working Group. 2014. Research misconduct involving noncompliance in human subjects research supported by the Public Health Service: Reconciling separate regulatory systems. *Hastings Center Report* 44(4): S2-S26.

BioMed Central. 2017. Publication Ethics. Available at: https://www.biomedcentral.com/getpublished/writing-resources/publication-ethics.

Bonito, A. J., C. Eicheldinger, N. Lenfestey, B. Munoz, D. Plotner, and F. Yu. 2009. *The Research Integrity Officer (RIO) Study*. Report submitted to ORI. Available at ori.hhs.gov/images/ddblock/RIOstudy.pdf. Accessed December 17, 2015.

Bonito, A. J., C. Eicheldinger, A. M. Greene, N. Lenfestey, J. Li, B. Munoz, D. Plotner, and F. Yu. 2010. *Evaluating the Effectiveness of Institutional Efforts to Educate Staff on Their Policies for Dealing with Research Misconduct and Research Integrity. Report to ORI*. Available at: ori.hhs.gov/images/ddblock/EvaluatingEffectivenessInstituions.pdf. Accessed December 17, 2015.

Bornemann-Cimenti, H., I. S. Szilagyi, and A. Sandner-Kiesling. 2015. Perpetuation of retracted publications using the example of the Scott S. Reuben case: Incidences, reasons and possible improvements. *Science and Engineering Ethics* 22(4): 1063-1072.

Bosch, X., and J. Ross. 2012. Ghostwriting: Research misconduct, plagiarism, or fool's gold? *American Journal of Medicine* 125(4): 324-326.

Bouri, S., M. J. Shun-Shin, G. D. Cole, J. Mayet, and D. P. Francis. 2013. Meta-analysis of secure randomised controlled trials of β-blockade to prevent perioperative death in non-cardiac surgery. *Heart* 100: 456-464.

BPC (Bipartisan Policy Center). 2009. *Improving the Use of Science in Regulatory Policy: Final Report*. Available at: http://bipartisanpolicy.org/wp-content/uploads/sites/default/files/BPC%20Science%20Report%20fnl.pdf.

Brand, A., L. Allen, M. Altman, M. Hlava, and J. Scott. 2015. Beyond authorship: attribution, contribution, collaboration, and credit. *Learned Publishing* 28: 151-155.

Braunschweiger, P., and K. W. Goodman. 2007. The CITI program: An international online resource for education in human subjects protection and the responsible conduct of research. *Academic Medicine* 82(9): 861-864.

Brent, E., and C. Atkisson. 2011. Accounting for cheating: An evolving theory and emergent themes. *Research in Higher Education* 52: 640-658.

Broad, W., and N. Wade. 1983. *Betrayers of the Truth*. New York: Simon & Schuster.

Brummel, B. J., C. K. Gunsalus, K. L. Anderson, and M. C. Loui. 2010. Development of role-play scenarios for teaching responsible conduct of research. *Science and Engineering Ethics* 16: 573-589.

Buchkowsky, S. S., and P. J. Jewesson. 2004. Industry sponsorship and authorship of clinical trends over 20 years. *Annals of Pharmacotherapy* 38(4): 579-585.

Buckheit, J. B., and D. L. Donoho. 1995. WaveLab and reproducible research. Pp. 55-81 in *Wavelets and Statistics,* A. Antoniadis and G. Oppenheim, eds. New York: Springer.

Bunton, S. A., and W. T. Mallon. 2007. The continued evolution of faculty appointment and tenure policies at U.S. medical schools. *Academic Medicine* 82: 281-289.

Butler, D. 2010. Journals step up plagiarism policing. *Nature* 466(7303): 167.

Butler, D. 2012. Hyped GM maize study faces growing scrutiny. *Nature* 490: 158.

Buyse, M., S. L. George, S. Evans, N. L. Geller, J. Ranstam, B. Scherrer E. Lesaffre, G. Murray, L. Edler, J. Hutton, T. Colton, P. Lachenbruch, and B. L. Verma. 1999. The role of biostatistics in the prevention, detection and treatment of fraud in clinical trials. *Statistics in Medicine* 18(24): 3435-3451.

Buzzelli, D. E. 1993. The definition of misconduct in science: A view from NSF. *Science* 259(5095): 584-585.

Cain, D. M., G. Loewenstein, and D. A. Moore. 2005. The dirt on coming clean: Perverse effects of disclosing conflicts of interest. *Journal of Legal Studies* 34(1): 1-25.

Cain, D. M., G. Loewenstein, and D. A. Moore. 2011. When sunlight fails to disinfect: Understanding the perverse effects of disclosing conflicts of interest. *Journal of Consumer Research* 37(5): 836-857.

Canary, H. E., and J. R. Herkert. 2012. Assessing ethics education in programs and centers: Challenges and strategies. Pp 38-43 *in Practical Guidance on Science and Engineering Ethics Education for Instructors and Administrators*. Washington, DC: The National Academies Press.

Carey, B. 2015. Science, now under scrutiny itself. *New York Times*. June 15. Available at: https://www.nytimes.com/2015/06/16/science/retractions-coming-out-from-under-science-rug.html.

Carpenter, S. 2012. Government sanctions Harvard psychologist. *Science* 337(610): 1283.

Cartlidge, E. 2012. Prison terms for L'Aquila experts shock scientists. *Science* 338(6106): 451-452.

Cartlidge, E. 2014. Updated: Appeals court overturns manslaughter convictions of six earthquake scientists. *Science Insider.* Available at: http://www.sciencemag.org/news/2014/11/updated-appeals-court-overturns-manslaughter-convictions-six-earthquake-scientists.

Casadevall, A., and F. C. Fang. 2012. Reforming science: Methodological and cultural reforms. *Infection and Immunity* 80(3): 891-896.

CASRAI. 2016. Project CRediT. Available at: http://docs.casrai.org/CRediT.

Castelvecchi, D. 2015. Physics paper sets record with more than 5,000 authors. *Nature*. May 15. Available at: http://www.nature.com/news/physics-paper-sets-record-with-more-than-5-000-authors-1.17567.

CCA (Council of Canadian Academies). 2010. *Honesty, Accountability and Trust: Fostering Research Integrity in Canada*. Ottawa. Available at: http://www.scienceadvice.ca/en/assessments/completed/research-integrity.aspx.

CGS (Council of Graduate Schools). 2012. *Research and Scholarly Education in Graduate Education*. Washington, DC: CGS.

Chalmers, M. 2015. Forsaken pentaquark particle spotted at CERN. *Nature* 523(7560): 267-268.

Chen, R., N. R. Desai, J. S. Ross, W. Zhang, K. H. Chau, and B. Wayda. 2016. Publication and reporting of clinical trial results: Cross sectional analysis across academic medical centers. *BMJ* 352: i637.

Chopra, V., and K. A. Eagle. 2012. Perioperative mischief: The price of academic misconduct. *American Journal of Medicine* 125(10): 953-955.

CHPS Consulting. 2000. *Analysis of Institutional Policies for Responding to Allegations of Scientific Misconduct*. Report submitted to Office of Research Integrity. Available at: ori.hhs.gov/sites/default/files/institutional_policies.pdf. Accessed December 17, 2015.

Clarke, R. V. 1995. Situational crime prevention. *Crime and Justice* 19: 91-150.

Clarkeburn, H., J. R. Downie, and B. Matthew. 2002. Impact of an ethics programme in a life sciences curriculum. *Teaching in Higher Education* 7: 65-79.

Clayton, S., and S. Opotow (eds.). 2003. *Identity and the Natural Environment*. Cambridge, MA: MIT Press.

Cloward, R. A., and L. E. Ohlin. 1960. *Delinquency and Opportunity*. New York: Free Press.

Cohen, A. K. 1965. The sociology of the deviant act: Anomie theory and beyond. *American Sociological Review* 30: 5-14.

Collins, F. 2010. Opportunities for research and NIH. *Science* 327(5961): 36.

Colquitt, J. A., D. E. Conlon, M. J. Wesson, C. O. L. H. Porter, and K. Y. Ng. 2001. Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology* 86(3): 425-445.

Commission on Research Integrity. 1995. *Integrity and Misconduct in Research*. Washington, DC: U.S. Department of Health and Human Services, Public Health Service.

COPE (Committee on Publication Ethics). 2011. *Code of Conduct and Best Practice Guidelines for Journal Editors*. Available at: http://publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf. Accessed July 23, 2013.

COPE. 2017. eLearning course on Publication Ethics. Available at: https://publicationethics.org/resources/e-learning.

Cope, M. B., and D. B. Allison. 2010. White hat bias: A threat to the integrity of scientific reporting. *Acta Paediatrica* 99: 1615-1617.

COS (Center for Open Science). 2015. Transparency and Openness Promotion (TOP) Guidelines. Available at: https://cos.io/top/#signatories.

Cossins, D. 2013. Settlement reached in misconduct case. *The Scientist*. April 19. Available at: http://www.the-scientist.com/?articles.view/articleNo/35216/title/Settlement-Reached-in-Misconduct-Case/.

Couzin, J., and K. Unger. 2006. Cleaning up the paper trail. *Science* 312: 38-43.

Couzin-Frankel, J. 2013. Fraudster blog author outed and comes clean. *Science*. Available at: http://news.sciencemag.org/2013/01/fraudster-blog-author-outed-and-comes-clean.

Crain, A. L., B. C. Martinson, and C. R. Thrush. 2013. Relationships between the Survey of Organizational Research Climate (SORC) and self-reported research practices. *Science and Engineering Ethics* 19(3): 835-850.

Cressey, D. 2013. "Rehab" helps errant researchers return to the lab. *Nature* 493(7431): 147.

CSE (Council of Science Editors). 2012a. CSE Publication Certificate Program. Available at https://www.councilscienceeditors.org/resource-library/cse-publication-certificate-program/.

CSE. 2012b. CSE's White Paper on Promoting Integrity in Scientific Journal Publications. Available at: http://www.councilscienceeditors.org/i4a/pages/index.cfm?pageid=3331. Accessed August 6, 2013.

Cyranoski, D. 2005. Stem-cell pioneer accused of faking data. *Nature*. December 15. doi:10.1038/news051212-14.

Cyranoski, D. 2014a. Stem-cell pioneer blamed media 'bashing' in suicide note. *Nature*. August 13. Available at: http://www.nature.com/news/stem-cell-pioneer-blamed-media-bashing-in-suicide-note-1.15715.

Cyranoski, D. 2014b. Stem-cell method faces fresh questions. *Nature* 507(7492): 283.

Cyranoski, D. 2014c. Stem-cell scientist found guilty of misconduct. *Nature News*. April 1. Available at: http://www.nature.com/news/stem-cell-scientist-found-guilty-of-misconduct-1.14974.

Davis, G. 2005. Doctors without orders. *American Scientist* 93(3, Suppl.). Available at: http://postdoc.sigmaxi.org/results/.

Davis, M., and A. Feinerman. 2010. Assessing graduate student progress in engineering ethics. *Science and Engineering Ethics* 18: 351-367.

Davis, M., and K. Riley. 2008. Ethics across the graduate engineering curriculum: An experiment in teaching and assessment. *Teaching Ethics* 9(1): 25-42.

Davis, M. S., M. Riske-Morris, and S. R. Diaz. 2007. Causal factors implicated in research misconduct: Evidence from ORI case files. *Science and Engineering Ethics* 13: 395-414.

DCSD (Danish Committees on Scientific Dishonesty). 2002. *Annual Report 2001*. Available at: ufm.dk/en/publications/2002/files-2002/annual-report-2001-the-danish-committees-on-scientific-d.pdf. Accessed April 19, 2015.

DCSD. 2009. *Guidelines for Good Scientific Practice*. Copenhagen.

DCSD. 2014. *The Danish Code of Conduct for Research Integrity*. Copenhagen.

DCSD. 2015. DCSD Website. Available at: ufm.dk/en/research-and-innovation/councils-and-commissions/the-danish-committees-on-scientific-dishonesty. Accessed April 19, 2015.

Derrick, E. G., H. J. Falk-Krzesinski, and M. R. Roberts (eds.). 2012. *Facilitating Interdisciplinary Research and Education: A Practical Guide*. Washington, DC: AAAS. Available at http://www.aaas.org/cspsp/interdisciplinary/guide.

Devereaux, M. L. 2014. Rethinking the meaning of ethics in RCR education. *Journal of Microbiology Education* 15(2): 165-168.

Ding, W. W., S. G. Levin, P. E. Stephan, and A. E. Winkler. 2010. The impact of information technology on academic scientists' productivity and collaboration patterns. *Management Science* 56(9): 1439-1461.

DOD (Department of Defense). 2004. Research integrity and Misconduct. Instruction No. 3210.7. Available at: http://www.dtic.mil/whs/directives/corres/pdf/321007p.pdf.

DOE-OIG (Department of Energy Office of Inspector General). 2014. Audit Report: Office of Science's Management of Research Misconduct Allegations. Available at: https://energy.gov/sites/prod/files/2014/08/f18/OAS-M-14-09_0.pdf.

DOI (Department of the Interior). 2010. Evaluation Report: Interior Lacks a Scientific Integrity Policy. Report No. WR-EV-MOA-0014-2009. Available at: https://www.doioig.gov/sites/doioig.gov/files/ScientificIntegrityPolicy.pdf.

DOI. 2014. Integrity of scientific and scholarly activities. Chapter 3 in *Department of the Interior Departmental Manual*: *Departmental Science Efforts*. Available at: http://elips.doi.gov/ELIPS/DocView.aspx?id=4056.

DOI. 2015. Integrity of Scientific and Scholarly Activities: Closed Cases. Available at: https://www.doi.gov/scientificintegrity/closed-cases.

DOI-OIG (Department of the Interior, Office of Inspector General). 2016. Inspection of Scientific Integrity Incident at USGS Energy Geochemistry Laboratory. Available at: https://www.doioig.gov/sites/doioig.gov/files/2016EAU010Public.pdf. Accessed March 16, 2016.

Donoho, D. 2012. Digitization and New Scientific Methods: Implications for Research Integrity. Presentation to the committee, Palo Alto, CA, July 9.

Donoho, D., A. Maleki, I. Rahman, M. Shahram, and V. Stodden. 2008. 15 Years of Reproducible Research in Computational Harmonic Analysis. Available at: https://statweb.stanford.edu/~donoho/Reports/2008/15YrsReproResch-20080426.pdf.

Dörner, D. 1996. *The Logic of Failure: Recognizing and Avoiding Error in Complex Situations*. New York: Basic Books.

Dresser, R. 1993. Sanctions for research misconduct: A legal perspective. *Academic Medicine* 68: S39-S43.

DuBois, J. M., J. M. Dueker, E. E. Anderson, and J. Campbell. 2008. The development and assessment of an NIH-funded research ethics training program. *Academic Medicine* 83(6): 596-603.

DuBois, J. M., E. E. Anderson, and J. Chibnall. 2013a. Assessing the need for a research ethics remediation program. *Clinical Translational Science* 6(3): 209-213.

DuBois, J. M., E. E. Anderson, J. Chibnall, K. Carroll, T. Gibb, C. Ogbuka, and T. Rubbelke. 2013b. Understanding research misconduct: A comparative analysis of 120 cases of professional wrongdoing. *Accountability in Research: Policies and Quality Assurance* 20(5-6): 320-338.

DuBois, J. M., J. T. Chibnall, R. Tait, and J. Vander Wal. 2016. Misconduct: Lessons from researcher rehab. *Nature* 534:175.

DUO (Education Executive Agency, Ministry of Education of The Netherlands). 2013. *Code of Conduct with Respect To International Student in Dutch Higher Education*. Available at: https://www.nuffic.nl/en/study-and-work-in-holland/code-of-conduct.

DVA (Department of Veterans Affairs). 2014. *VHA Handbook 1058.02: Research Misconduct*.. Available at: http://www.va.gov/vhapublications/ViewPublication.asp?pub_ID=2982.

DVA. 2015. Congressional Submission FY 2016 Funding and FY 2017 Advance Appropriations. Available at: http://www.va.gov/budget/docs/summary/Fy2016-VolumeII-MedicalProgramsAndInformationTechnology.pdf.

Dyer, C. 2011. The fraud squad. *BMJ* 342(June); doi: https://doi.org/10.1136/bmj.d4017.

*Economist, The*. 2011. An array of errors. September 10. Available at: http://www.economist.com/node/21528593.

Edsall, J. 1975. Scientific freedom and responsibility. *Science* 188(4189): 687-693.

Elliott, D., and J. Stern. 1996. Evaluating teaching and students' learning of academic research ethics. *Science and Engineering Ethics* 2(3): 345-366.

Elsevier. 2017. Publishing Ethics Resource Kit (PERK) for Editors. Available at: https://www.elsevier.com/editors/perk.

Enserink, M. 2012. Final report on Stapel also blames field as a whole. *Science* 338: 1270-1271.

Enserink, M. 2014. Sabotaged scientist sues Yale and her lab chief. *Science* 343: 1065-1066.

EPA (Environmental Protection Agency). 2006. Policy and Procedures for Addressing Research Misconduct. EPA Order 3120.5. Available at: http://ori.hhs.gov/sites/default/files/epapolicy.pdf.

EPA. 2012. Scientific Integrity Policy. Available at: http://www.epa.gov/sites/production/files/2014-02/documents/scientific_integrity_policy_2012.pdf.

Epskamp, S., and M. B. Nuijten. 2016. Statcheck: Extract statistics from articles and recompute *p* values. Available at: http://CRAN.Rproject.org/package=statcheck (R package version 1.2.2).

EQUATOR (*Enhancing the QUAlity and Transparency of health Research)*http://www.equator-network.org/.

ESF-ALLEA (European Science Foundation and All European Academies). 2011. *The European Code of Conduct for Research Integrity*. Available at: www.esf.org/fileadmin/Public_documents/Publications/Code_Conduct_ResearchIntegrity.pdf. Accessed April 18, 2015.

Evetts, J. 2013. Professionalism: Value and ideology. *Current Sociology* 61(5-6): 778-796.

Fanelli, D. 2009. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLOS ONE* 4(5): e5738.

Fanelli, D. 2010. Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLOS ONE* 5(4): e10271; doi:10.1371/journal.pone.0010271.

Fanelli, D. 2012. Negative results are disappearing from most disciplines and countries. *Scientometrics* 90: 891-904.

Fang, F. C., R. G. Steen, and A. Casadevall. 2012. Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences* 109(42): 17028-17033.

Feeley, J. 2012. Pfizer paid $896 million in Prempro settlements. *Bloomberg Business Week*. June 19.

Ferrer-Negron, J., W. Frey, E. O'Neill-Carrillo, D. Valdes, and C. Rios-Valazquez. 2009. *Graduate Education in Research Ethics for Scientists and Engineers: Final Report*. Available at: http://scholarworks.umass.edu/esence/376.

Fienberg, S., M. E. Martin, and M. L. Straff (eds.). 1985. *Sharing Research Data*. Washington, DC: National Academy Press.

Fischer, B. A., and M. J. Zigmond. 2001. Promoting responsible conduct in research through "survival skills" workshops: Some mentoring is best done in a crowd. *Science and Engineering Ethics* 7(4): 563-587.

Fischman, J. 2012. Fake peer reviews, the latest form of scientific fraud, fool journals. *Chronicle of Higher Education*. September 30.

Fitch, A. 2004. Public and private ethics in service learning for chemistry. *Abstracts of Papers of the American Chemical Society* 227: U667-U667.

Folger, R. G., and R. Cropanzano. 1998. *Organizational Justice and Human Resource Management*. Thousand Oaks, CA: SAGE.

Frankel, M. S., and S. J. Bird. 2003. The role of scientific societies in promoting research integrity. *Science and Engineering Ethics* 9(2): 139-140.

Freeman, R. B., and A. M. Gelber. 2006. *Optimal Inequality/Optimal Incentives: Evidence from a Tournament*. National Bureau of Economic Research. Available at: http://www.nber.org/papers/w12588. Accessed October 10, 2012.

Freeman, R. B., E. Weinstein, E. Marincola, J. Rosenbaum, and F. Solomon 2001a. *Careers and Rewards in Bio Sciences: The Disconnect Between Scientific Progress and Career Progression*. Bethesda, MD: American Society for Cell Biology.

Freeman, R., E. Weinstein, E. Marincola, J. Rosenbaum, and F. Solomon 2001b. Competition and careers in biosciences. *Science* 294: 2293-2294.

Freedman, L. P., I. M. Cockburn, and T. S. Simcoe. 2015. The economics of reproducibility in preclinical research. *PLOS Biology* 13(6); doi.org/10.1371/journal.pbio.1002165.

Frische, S. 2012. It is time for full disclosure of author contributions. *Nature* 489(7417): 475.

Frugoli, J. 2002. Using "ethics labs" to set a framework for ethical discussion in an undergraduate science course. Presented at 11th Annual Meeting of Association for Practical and Professional Ethics, Cinncinati, OH.

Fugh-Berman, A. J. 2010. The haunting of medical journals: How ghostwriting sold "HRT." *PLOS Medicine* 7(9): e1000335.

Galipeau, J., D. Moher, B. Skidmore, C. Campbell, P. Hendry, D. W. Cameron, P. C. Hébert, and A. Palepu. 2013. Systematic review of the effectiveness of training programs in writing for scholarly publication, journal editing, and manuscript peer review (protocol). *Systematic Reviews* 2(41). June. Available at: http://www.systematicreviewsjournal.com/content/2/1/41#B1. Accessed August 12, 2013.

Gardner, W. 2006. The social cost of scientific misconduct and questionable research practices. Presentation at the ORI Research on Research Integrity (RRI) Conference.

Garfinkel, S. 2012. Handling Cases: From Allegations to Closure, Problems to Avoid. Presentation to the committee. August 14.

GBSI (Global Biological Standards Institute). 2015. *The Case for Standards in Life Science Research*. Washington, DC: GBSI.

Gino, F., and D. Ariely. 2011. The Dark Side of Creativity: Original Thinkers Can Be More Dishonest. Harvard Business School NOM Unit Working Paper No. 11-064. January 11.

Gino, F., S. Ayal, and D. Ariely. 2009. Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological Science* 20(3): 393-398.

Gino, F., M. Schweitzer, N. Mead, and D. Ariely. 2011. Unable to resist temptation: How self-control depletion promotes unethical behavior. *Organizational Behavior and Human Decision Processes* 115(2): 191-203.

Glazer, S. 1997. Combating scientific misconduct. *CQ Researcher* 7: 1-24. Retrieved from http://library.cqpress.com/.

Gluckman, P. 2014. Policy: The art of science advice to government. *Nature* 507(7491): 163-165.

Godfrey, B. B. 2007. *Investigation of Alleged Research Misconduct by Lincoln Laboratory Members of 1998-5 POET Study Team*. Available at: https://orgchart.mit.edu/sites/default/files/ltc/pdf/InvestigationReport01292007.pdf.

Godlee, F., J. Smith, and H. Marcovitch.. 2011. Wakefield's article linking MMR vaccine and autism was fraudulent. *BMJ* (Jan. 5): 342. Available at: http://www.bmj.com/content/342/bmj.c7452. Accessed November 12, 2016.

Gold, B. D. 1993. Congressional activities regarding misconduct and integrity in science. In *Responsible Science, Vol. II*. Washington, DC: National Academy Press.

Goldenring, James R. 2010. Perspective: Innocence and due diligence: managing unfounded allegations of scientific misconduct. Academic medicine 85: 527-530.

Goodman, S. 2016. Robert De Niro pulls anti-vaccine documentary from Tribeca Film Festival. *New York Times*. March 26.

Goodstein, D. 2010. *On Fact and Fraud*. Princeton, NJ: Princeton University Press.

Goodwin, M., A. Bonham, A. Mazzaschi, H. Alexander, and J. Krakower. 2011. Sponsored program salary support to medical school faculty in 2009. *AAMC Analysis in Brief* 11(1). Available at: https://www.aamc.org/download/170836/data/aibvol11_no1.pdf.

Gorman, E. H., and R. L. Sandefur. 2011. "Golden Age," quiescence, and revival: How the sociology of professions became the study of knowledge-based work. *Work and Occupations* 38(3): 275-302.

Graf, C., L. Deakin, M. Docking, J. Jones, S. Joshua, T. McKerahan, M. Ottmar, A. Stevens, E. Wates, and D. Wyatt. 2015. Essay on *Best Practice Guidelines on Publishing Ethics: A Publisher's Perspective*, 2nd ed. *Advanced Materials* 27: 370-387.

GRC (Global Research Council). 2013. *Statement of Principles on Research Integrity*. Available at: http://grc.s2nmedia.com/statement-principles-research-integrity. Accessed August 15, 2013.

Greenberg, D. F. 1988. *The Construction of Homosexuality*. Chicago: University of Chicago Press.

Greene, A. M., A. J. Bonito, J. Amoozegar, C. Eicheldinger, N. Lenfestey, and D. Plotner. 2011. *Whistleblowers in Research Misconduct Cases: RIOs Preparation of Whistleblowers and Their Reports on Questions Whistleblowers Raised*. Report Submitted to ORI. Available at: https://ori.hhs.gov/images/ddblock/Whistleblower.pdf. Accessed December 17, 2015.

Greenland, P., and P. B. Fontanarosa. 2012. Ending honorary authorship. *Science* 337(6098): 1019.

Grens, K. 2013a. Rampant plagiarism in two journals. *The Scientist*. May 22.

Grens, K. 2013b. What to do about "Clare Francis." *The Scientist*. September 14.

Grieneisen, M. L., and M. Zhang. 2012. A comprehensive survey of retracted articles from the scholarly literature. *PLOS ONE 7*(10): e44118.

Grifo, F. 2013. *Federal Agency Scientific Integrity Policies*. Union of Concerned Scientists. http://www.ucsusa.org/our-work/center-science-and-democracy/promoting-scientific-integrity/federal-agency-si-policies.html#.VkNSn3arSUl.

Grinnell, F. 2014. The interrelationship between research integrity, conflict of interest, and the research environment. *Journal of Microbiology & Biology Education* 15(2): 162-164.

Gunsalus, C. K. 1993 Institutional structure to ensure research integrity. *Academic Medicine* 68(9): S33-S38.

Gunsalus, C. K. 1997. Rethinking Unscientific Attitudes About Scientific Misconduct. *Chronicle of Higher Education*. March 28.

Gunsalus, C. K. 1998a. How to blow the whistle and still have a career afterwards. *Science and Engineering Ethics* 4(1): 51-64.

Gunsalus, C. K. 1998b. Preventing the need for whistleblowing: Practical advice for university administrators. *Science and Engineering Ethics* 4(1): 75-93.

Gustafsson, B., G. Hermerén, and B. Petersson. 2006. *Good Research Practice—What Is It?* Stockholm: Swedish Research Council.

GUSTO Investigators. 1993. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *New England Journal of Medicine* 329: 673-682.

Hackett, E. J. 1994. A social control perspective on scientific misconduct. *Journal of Higher Education* 65(3): 242-260.

Hames, I. 2013. *COPE Ethical Guidelines for Peer Reviewers*. COPE Council. Available at: http://publicationethics.org/files/Ethical_guidelines_for_peer_reviewers_0.pdf.

Hanson, B., A. Sugdon, and B. Alberts. 2011. Making data maximally available. *Science* 331(6018):649.

Harris, C. E. 2008. The good engineer: Giving virtue its due in engineering ethics. *Science and Engineering Ethics* 13: 153-164.

Harvard Medical School. 1999. Authorship Guidelines. Available at: https://hms.harvard.edu/sites/default/files/assets/Sites/Ombuds/files/AUTHORSHIP%20GUIDELINES.pdf.

Heitman, E., and S. Litewka. 2011. International perspectives on plagiarism and considerations for teaching international trainees. *Urologic Oncology* 29(1): 104-108.

Heitman, E., and J. Petty. 2010. Preparing students to navigate cross-national differences in the research environment: The role of research integrity education. Pp. 201-214 in *International Research Collaborations: Much to Be Gained, Many Ways to Get in Trouble,* M. S. Anderson and N. H. Steneck, eds. New York: Routledge.

Heitman, E., C. H. Olsen, L. Anestidou, and R. E. Bulger. 2007. New graduate students' baseline knowledge of the responsible conduct of research. *Academic Medicine* 82(9): 838-845.

Hellweg, B. 2015. At med school, grants provide uncertain salary. *Yale Daily News,* April 9.

HHMI (Howard Hughes Medical Institute). 2007. Research Policies: Scientific Misconduct. Available at www.hhmi.org/sites/default/files/About/Policies/sc_200.pdf. Accessed April 19, 2015.

HHS (Department of Health and Human Services). 2000. Public Health Service standards for the protection of research misconduct whistleblowers. *Federal Register* 65: 70830-70841.

HHS. 2005. Title 42: Public Health Service (PHS). Part 93: Policies on Research Misconduct. Code of Federal Regulations. June.

HHS. 2009. Title 45: Public welfare. Part 46: Protection of human subjects. Code of Federal Regulations. Available at: http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html.

HHS. 2011a. ANPRM for Revision to Common Rule. Available at: http://www.hhs.gov/ohrp/humansubjects/anprm2011page.html.

HHS. 2011b. Responsibility of applicants for promoting objectivity in research for which public health service funding is sought and responsible prospective contractors; final rule. *Federal Register* 76(165): 53256-53293. Available at: https://www.gpo.gov/fdsys/pkg/FR-2011-08-25/pdf/2011-21633.pdf.

HHS. 2012. Food and Drug Administration disqualification of a clinical investigator. *Federal Register* 77(83): 25353-25361.

HHS. 2014a. Findings of research misconduct. Notice No. NOT-OD-11-070. Available at: http://grants.nih.gov/grants/guide/notice-files/NOT-OD-11-070.html.

HHS. 2014b. Fiscal 2015 Budget Justification. Available at: http://www.hhs.gov/sites/default/files/budget/fy2015/fy-2015-hhs-congressional-budget-justification.pdf, Accessed December 20, 2015.

HHS. 2016. NPRM for Revisions to the Common Rule: HHS Announces Proposal to Improve Rules Protecting Human Research Subjects. NPRM 2015—Summary. Available at: http://www.hhs.gov/ohrp/index.html.

Hirschi, T. 1969. *Causes of Delinquency*. Berkeley: University of California Press.

Holdren, J. 2010. Scientific Integrity. Memorandum for the Heads of Executive Departments and Agencies. December 17. Executive Office of the President, Office of Science and Technology Policy. Available at https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/scientific-integrity-memo-12172010.pdf.

Holdren, J. P. 2013. Increasing Access to the Results of Federally Funded Scientific Research. Memorandum for the Heads of Executive Departments and Agencies. February 22. Executive Office of the President, Office of Science and Technology Policy. Available at: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Hornyak, T. 2012. Did Higgs yield the most authors in a science study? *CNET*. September 10. Available at: http://news.cnet.com/8301-17938_105-57509551-1/did-higgs-yield-the-most-authors-in-a-science-study/. Accessed March 28, 2013.

Hutchins, W. J. 1982. The evolution of machine translation systems. In *Practical Experience of Machine Translation,* V. Lawson, ed. Amsterdam: North-Holland. Available at: http://www.mt-archive.info/Aslib-1981-Hutchins-1.pdf.

IAC (InterAcademy Council). 2010. *Climate Change Assessments: Review of the Processes and Procedures of the IPCC*. Amsterdam: IAC.

IAC-IAP (InterAcademy Council and InterAcademy Panel). 2012. *Responsible Conduct in the Global Research Enterprise: A Policy Report*. Amsterdam: IAC-IAP.

IAP (InterAcademy Partnership). 2016. *Doing Global Science: A Guide to Responsible Conduct in the Global Research Enterprise*. Princeton, NJ: Princeton University Press.

IAP (InterAcademy Panel on International Issues). 2005. *IAP Statement on Biosecurity*. Available at: http://www.interacademies.net/File.aspx?id=5401.

ICB (Irish Council for Bioethics). 2010. *Recommendations for Promoting Research Integrity*. Dublin: ICB.

ICMJE (International Committee of Medical Journal Editors). 2013a. Defining the Role of Authors and Contributors. Available at: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html.

ICMJE. 2013b. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Ethical Considerations in the Conduct and Reporting of Research: Conflicts of Interest. Available at: http://www.icmje.org/ethical_4conflicts.html. Accessed July 23, 2013.

Ioannidis, J. P. A. 2005. Why most published research findings are false. *PLOS Medicine* 2(8): e124.

Ioannidis, J. P. A. 2012. Why science is not necessarily self-correcting. *Perspectives on Psychological Science* 7(6): 645-654.

IOM (Institute of Medicine). 2009. *Conflict of Interest in Medical Research, Education, and Practice*. Washington, DC: The National Academies Press.

IOM. 2015. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risks*. Washington, DC: The National Academies Press.

IOM-NRC (Institute of Medicine and National Research Council). 2002. *Integrity in Scientific Research: Creating an Environment That Promotes Responsible Conduct*. Washington, DC: National Academies Press.

Iqbal, N., W. A. Bhatti, and A. Zaheer. 2013. Corporate Ethical values, group creativity, job commitment and performance: The effect of work response on work context. *Management and Marketing* 8: 1. January. Available at: http://www.managementmarketing.ro/pdf/articole/302.pdf.

Ishii, S., A. Iwama, H. Koseki, Y. Shinkai, T. Taga, and J. Watanabe. 2014. *Report on STAP Cell Research Paper Investigation*. RIKEN. Available at: www3.riken.jp/stap/e/f1document1.pdf. Accessed December 16, 2015.

Iverson, M., M. S. Frankel, and S. Siang. 2003. Scientific societies and research integrity: What are they doing and how well are they doing it? *Science and Engineering Ethics* 9(2): 141-158.

James, W. H. 1995. Fraud and hoaxes in science. *Nature* 377: 474.

Jaschik, S. 2015. Who gets credit? *Inside Higher Ed*. August 24. Available at: https://www.insidehighered.com/news/2015/08/24/research-reveals-significant-share-scholarly-papers-have-guest-or-ghost-authors.

Jayaraman, K. S. 1994. Fossil inquiry finds Indian geologist guilty of plagiarism. *Nature* 369: 698.

Jha, A. 2012. False positives: Fraud and misconduct are threatening scientific research. *The Guardian*. September 13. Available at: http://www.theguardian.com/science/2012/sep/13/scientific-research-fraud-bad-practice.

Jogalekar, A. S. 2015. Social media, peer review, and responsible conduct of research (RCR) in chemistry: Trends, pitfalls, and promises. *Accountability in Research: Policies and Quality Assurance* 22(6): 402-430.

John, L. K., G. Loewenstein, and D. Prelac. 2012. Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science* 23(5): 524-532.

Joppa, L. N., G. McInerny, R. Harper, L. Salido, K. Takeda, K. O'Hara, D. Gavaghan, and S. Emmott. 2013. Troubling trends in scientific software use. *Science* 340(6134): 814-815.

Jordan, T. 2013. Lessons of L'Aquila for operational earthquake forecasting. *Seismological Research Letters* 84(1): 4-7.

JSA (Japan Society of Anesthesiologists). 2012. *The Results of Investigation into Dr.Yoshitaka Fujii's papers*. Available at: http://www.anesth.or.jp/english/pdf/news20120629.pdf (accessed January 13, 1012).

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, D. 2012. A proposal to deal with questions about priming effects. Open Letter. Available at: http://www.nature.com/polopoly_fs/7.6716.1349271308!/suppinfoFile/Kahneman%20Letter.pdf.

Kahneman, D., and A. Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47(2): 263-291.

Kaiser, J. 1997. The "Gallo case": Popovic strikes back. *Science* 275(5302): 920-921.

Kalb, P. E., and K. G. Koehler. 2002. Legal issues in scientific research. *JAMA* 287(1): 85-91.

Kalichman, M. W. 2012. Why teach research ethics? Pp 5-16 in *Practical Guidance on Science and Engineering Ethics Education for Instructors and Administrators*. Washington, DC: The National Academies Press.

Kalichman, M. W., and P. J. Friedman. 1992. A pilot study of biomedical trainees' perceptions concerning research ethics. *Academic Medicine* 67(11): 769-775.

Kalichman, M. W., and D. K. Plemmons. 2007. Reported goals for responsible conduct of research courses. *Academic Medicine* 82(9): 846-852.

Kaufman, M. 2012. Journal retreats from controversial arsenic paper. *The Washington Post*. July 8.

Kevles, D. 1998. *The Baltimore Case: A Trial of Politics, Science and Character*. New York: W.W. Norton & Co.

Keystone Center. 2012. *Improving the Use of Science in Regulatory Decision-Making: Dealing with Conflict of Interest and Bias in Scientific Advisory Panels, and Improving Systematic Scientific Reviews*. September 18. Available at: https://www.keystone.org/images/keystone-center/spp-documents/Health/Research%20Integrity%20Rountable%20Report.pdf. Accessed August 15, 2013.

King, C. 2013. Single-author papers: A waning share of output, but still providing the tools for progress. *Science Watch*. September. Available at: sciencewatch.com/articles/single-author-papers-waning-share-output-still-providing-tools-progress. Accessed December 19, 2014.

KNAW (Royal Netherlands Academy of Arts and Sciences). 2013. *Responsible Research Data Management and the Prevention of Scientific Misconduct*. Amsterdam: KNAW.

Koch, C., and C. Schmidt. 2010. Disclosing conflicts of interest—Do experience and reputation matter? *Accounting, Organizations and Society* 35(1): 95-107.

Kornfeld, D. S. 2012. Research misconduct: The search for a remedy. *Academic Medicine* 87(7): 877-882.

Krohn, M. 1995. Control and deterrence theories of crime. In *Criminology: A Contemporary Handbook*. Belmont, CA: Wadsworth.

Kuhn, T. S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kunst, F. et al. 1997. The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*. *Nature* 390: 249-256.

Kuta, S. 2014. CU-Boulder scientists speak out on research misconduct claim. *Boulder Daily Camera*. February 7.

LaCour, M. J., and D. P. Green. 2014. When contact changes minds: An experiment on transmission of support for gay equality. *Science* 346(6215): 1366-1369. RETRACTED.

Landis, S. C., S. G. Amara, K. Asadullah, C. P. Austin, R. Blumenstein, E. W. Bradley, and S. D. Silberberg. 2012. A call for transparent reporting to optimize the predictive value of preclinical research. *Nature* 490(7419): 187-191.

Lang, J. M. 2013. *Cheating Lessons: Learning from Academic Dishonesty*. Cambridge, MA: Harvard University Press.

Leape, L. L. 2010. Transparency and Public Reporting Are Essential for a Safe Health Care System. The Commonwealth Fund. Available at: http://www.commonwealthfund.org/publications/perspectives-on-health-reform-briefs/2010/mar/transparency-and-public-reporting-are-essential-for-a-safe-health-care-system.

Lesser, L. I., C. B. Ebbeling, M. Goozner, D. Wypij. and D. S. Ludwig. 2007. Relationship between funding source and conclusion among nutrition-related scientific articles. *PLOS Medicine* 4(1): e5. doi: 10.1371/journal.pmed.0040005.

Levelt, W. J. M., P. Drenth, and E. Noort (eds.). 2012. *Flawed Science: The Fraudulent Research Practices of Social Psychologist Diederik Stapel*. Tilburg University, University of Amsterdam, and University of Groningen. available at: https://www.tilburguniversity.edu/upload/3ff904d7-547b-40ae-85fe-bea38e05a34a_Final%20report%20Flawed%20Science.pdf.

Levine, A. S., R. J. Alpern, N. C. Andrews, K. Antman, J. R. Balser, J. M. Berg, P. B. Davis, J. G. Fitz, R. N. Goldman, J. L. Jameson, V. S. Lee, K. S. Polonsky, M. S. Rappley, E. A. Reece, P. B. Rothman, D. A. Schwinn, L. J. Shapiro, and A. M. Spiegel. 2015. Research in academic medical centers: Two threats to sustainable support. *Science Translational Medicine* 7(289): 289fs22.

Liberman, M. 2012. Lessons for Responsible Science from DARPA's Programs in Human Language Technology, Presentation to the Committee. July 9.

Lock S., F. Wells, M. Farthing, and S. Evans. 2001. Statistical aspects of the detection of fraud. In *Fraud and Misconduct in Medical Research,* 3rd ed., S. Lock, F. Wells, and M. Farthing eds. London: BMJ.

Loewenstein, G., D. M. Cain, and S. Sah. 2011. The limits of transparency: Pitfalls and potential of disclosing conflicts of interest. *American Economic Review: Papers and Proceedings* 101(3): 423-428.

Long, T. C., M. Errami, A. C. George, Z. Sun, and H. R. Garner. 2009. Responding to possible plagiarism. *Science* 323(5919): 1293-1294.

Louis, K. S., M. S. Anderson, and L. Rosenberg. 1995. Academic misconduct and values: The department's influence. *Review of Higher Education* 18(4): 393-422.

Lowry, Paul Benjamin, Gregory D. Moody, and James Gaskin, et. al. Evaluating Journal Quality and the Association for Information Systems (AIS) Senior Scholars' Journal Basket Via Bibliometric Measures: Do Expert Journal Assessments Add Value? (2013). MIS Quarterly Vol. 37(4), 993–1012. Available at SSRN: https://ssrn.com/abstract=2186798

Lubalin, J., M.-A. E. Ardini, and J. Matheson, J.. 1995. *Consequences of Whistleblowing for the Whistleblower in Misconduct in Science Cases*. Prepared for the Office of Research Integrity. Washington, DC: Research Triangle Institute.

Macrina, F. L. 2007. Scientific societies and promotion of the responsible conduct of research: Codes, policies, and education. *Academic Medicine* 82(9): 865-869.

Maher, B. 2010. Research integrity: Sabotage! *Nature* 467: 516-518.

Mani, A., S. Mullainathan, E. Shafir, and J. Zhao. 2013. Poverty impedes cognitive function. *Science* 341: 976-980.

Maron, D. F. 2015. Spreading measles outbreak also takes heavy economic toll. *Scientific American*. February 5.

Martin, A. 2012. Building a showcase campus, using an I.O.U. *New York Times*. December 13.

Martinson, B. C., M. S. Anderson, and R. de Vries. 2005. Scientists behaving badly. *Nature* 435: 737-738.

Martinson, B. C., M. S. Anderson, A. L. Crain, and R. de Vries. 2006. Scientists' perceptions of organizational justice and self-reported misbehaviors. *Journal of Empirical Research on Human Research Ethics* 1(1): 51-66.

Martinson, B. C., A. L. Crain, M. S. Anderson, and R. de Vries. 2009. Institutions' expectations for researchers' self-funding, federal grant holding and private industry involvement: Manifold drivers of self-interest and researcher behavior. *Academic Medicine* 84: 1491-1499.

Martinson, B. C., A. L. Crain, R. de Vries, and M. S. Anderson. 2010. The importance of organizational justice in ensuring research integrity. *Journal of Empirical Research on Human Research Ethics* 5: 67-83.

Martinson, B. C., C. R. Thrush and A. L. Crain. 2013. Development and validation of the Survey of Organizational Research Climate (SORC). *Science and Engineering Ethics* 19(3): 813-834.

Marušić, A., L. Bošnjak. and A. Jerončić. 2011. A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *PLOS ONE* 6(9): e23477. Available at: http://www.plosone.org/article/info%253Adoi%252F10.1371%252Fjournal.pone.0023477. Accessed August 1, 2013.

Matza, D. 1964. *Delinquency and Drift*. New York: John Wiley & Sons.

Matza, D. 1969. *Becoming Deviant*. Englewood Cliffs, NJ: Prentice-Hall.

May, T. 2013. Embedding research integrity in graduate education—Michigan State University. Presentation at the World Conference on Research Integrity, Montreal. May 8.

Mazar, N., and D. Ariely. 2015. Dishonesty in scientific research. *Journal of Clinical Investigation* 125(11): 3993-3996.

Mazar, N., O. Amir, and D. Ariely. 2008. The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research* 45(6): 633-644.

McCook, A. 2015. How long does it take to retract a paper? A look at the Eric Poehlman record. Retraction Watch. Available at: http://retractionwatch.com/2015/09/09/how-long-does-it-take-to-retract-a-paper-a-look-at-the-eric-poehlman-record/.

McCormack, W. T., and C. W. Garvan. 2014. Team-based learning instruction for responsible conduct of research positively impacts ethical decision-making. *Accountability in Research* 21(1): 34-49.

McCormick, J. B., A. M. Boyce, J. M. Ladd, and M. K. Cho. 2012. Barriers to considering ethical and societal implications of research: Perceptions of life scientists. *AJOB Primary Research* 3(3): 40-50.

McKenzie, F. 2014. *Complex Adaptive Systems: Implications for Leaders, Organizations, Government and Citizens*. Australian Futures Project. http://australianfutures.org/wp-content/uploads/2016/01/AFP-PB1-Complex-Adaptive-Systems-140514.pdf.

McNaull, A. D. 2014. PCAST discusses data reproducibility issues. *AIP Bulletin 32* (March 5). Available at: http://www.aip.org/fyi/2014/pcast-discusses-data-reproducibility-issues. Accessed August 4, 2014.

McNutt, M. 2015. Expression of concern on LaCour and Green, *Science* 346(6215), 1366-1369. *Science* 348(6239): 1100.

Medawar P. 1996. *The Strange Case of the Spotted Mice and Other Classic Essays on Science*. Oxford: Oxford University Press.

Medeiros, K. E., J. T. Mecca, C. Gibson, V. D. Giorgini, M. D. Mumford, L. Devenport, and S. Connelly. 2014. Biases in ethical decision making among university faculty. *Accountability in Research* 21(4): 218-240.

Merton, R. K. 1938. Social structure and anomie. *American Sociological Review* 3: 672-682.

Merton, R. K. 1973. *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago: University of Chicago Press.

Mervis, J. 2013. NSF audit of successful proposals finds numerous cases of alleged plagiarism. *Science Insider*. Available at: http://www.sciencemag.org/news/2013/03/nsf-audit-successful-proposals-finds-numerous-cases-alleged-plagiarism.

Michalek, A. M., A. D. Hutson, C. P. Wicher, and D. L. Trump. 2010. The costs and underappreciated consequences of research misconduct: A case study. *PLOS Medicine* 7(8): e1000318; doi:10.1371/journal.pmed.1000318.

Mitroff, I. I. 1974. Norms and counter-norms in a select group of the Apollo moon scientists: A case study of the ambivalence of scientists. *American Sociological Review* 39(4): 579-595.

MPG (Max Planck Society). 2009. Rules of good scientific practice. Munich.

Mumford, M. D., S. T. Murphy, S. Connelly, J. H. Hill, A. L. Antes, R. P. Brown, and L. D. Devenport. 2007. Environmental influences on ethical decision making: Climate and environmental predictors of research integrity. *Ethics & Behavior* 17(4): 337-366.

Mumford, M. D., S. Connelly, R. P. Brown, S. T. Murphy, J. H. Hill, A. L. Antes, E. P. Waples, and L. D. Devenport. 2008. A sensemaking approach to ethics training for scientists: Preliminary evidence of training effectiveness. *Ethics & Behavior* 18(4): 315-339.

NAE (National Academy of Engineering). 2003-17. Online Ethics Center. Available from: http://www.onlineethics.org/about/18848.aspx.

NAE. 2009. *Ethics Education and Scientific and Engineering Research: What's Been Learned? What Should Be Done? Summary of a Workshop*. Washington, DC: The National Academies Press.

NAS (National Academy of Sciences). 1989. *On Being a Scientist*, 1st Ed. Washington, DC: National Academy Press.

NAS-NAE-IOM (National Academy of Sciences, National Academy of Engineering, and Institute of Medicine). 1992. *Responsible Science: Ensuring the Integrity of the Research Process*. Washington, DC: National Academy Press.

NAS-NAE-IOM. 1995. *On Being a Scientist: Responsible Conduct in Research*, 2nd Ed. Washington, DC: National Academy Press.

NAS-NAE-IOM. 2009a. *Ensuring the Integrity, Accessibility and Stewardship of Research Data*. Washington, DC: The National Academies Press.

NAS-NAE-IOM. 2009b. *On Being a Scientist: A Guide to Responsible Conduct in Research*, 3rd Ed. Washington, DC: The National Academies Press.

NAS-NAE-IOM. 2014. *The Postdoctoral Experience Revisited*. Washington, DC: The National Academies Press.

Nath, I., and E. L. Winnacker. 2012. Responsible research conduct. *Science* 338(6109): 863.

National Academies of Sciences, Engineering, and Medicine. 2016. *Optimizing The Nation's Investment in Academic Research: A New Regulatory Framework for the 21st Century*. Washington, DC: The National Academies Press.

National Institutes of Health Revitalization Act of 1993. 1993. Pub. L. No. 103-43, § 161, 107 Stat. 122, 140-41 (codified as amended at 42 U.S.C. § 289b (2000)).

*Nature*. 2005. Timeline of a controversy: A chronology of Woo Suk Hwang's stem-cell research. December 19. doi:10.1038/news051219-3.

*Nature*. 2009. Editorial: Authorship policies. *Nature* (458)1078. April 29. Available at: http://www.nature.com/nature/journal/v458/n7242/full/4581078a.html. Accessed August 1, 2013.

*Nature*. 2013. Authors and Referees: Editorial Policies: Availability of Data and Materials. Available at: http://www.nature.com/authors/policies/availability.html.

*Nature*. 2014. Retraction challenges. *Nature* 514(5). October 2. Available at: http://www.nature.com/news/retraction-challenges-1.16023.

*Nature*. 2015a. Announcement: Time to tackle cells' mistaken identity (Editorial). 520: 264. April 15.

*Nature*. 2015b. Challenges in Irreproducible Research. Article Archive. Available at: http://www.nature.com/nature/focus/reproducibility/index.html.

Neale, A. V., J. Northrup, R. Dailey, E. Marks, and J. Abrams. 2007. Correction and use of biomedical literature affected by scientific misconduct. *Science and Engineering Ethics* 13(1): 5-24. March.

*Neurology*. 2016. Information for Authors. Available at: m.neurology.org/site/misc/auth2.xhtml#AUTHORSHIP. Accessed June 16, 2016.

NHMRC-ARC-UA (National Health and Medical Research Council, Australian Research Council, and Universities Australia). 2007. *Australian Code for the Responsible Conduct of Research*. Canberra: Australian Government.

Nicholson, J. M., and J. P. A. Ioannidis. 2012. Research grants: Conform and be funded. *Nature* 492: 34-36. December 6.

Nielsen, M. 2012. *Reinventing Discovery: The New Era of Networked Science*. Princeton, NJ: Princeton University Press.

NIH (National Institutes of Health). 2009. Update on the Requirement for Instruction in the Responsible Conduct of Research. November 24. Available at: http://grants.nih.gov/grants/guide/notice-files/not-od-10-019.html.

NIH. 2011. Responsibility of applicants for promoting objectivity in research for which PHS funding sought (42 CFR Part 50 Subpart F). Available at: http://grants.nih.gov/grants/policy/coi/coi_faqs.htm#3152.

NIH. 2012a. *A Guide to Handling Research Misconduct Allegations*. Available at: http://sourcebook.od.nih.gov/ResEthicsCases/NIH%20Misconduct2.pdf. Accessed January 10, 2013.

NIH. 2012b. *Biomedical Research Workforce Working Group Report*. Bethesda, MD: NIH.

NIH. 2014. Update: New Biographical Sketch Format Required for NIH and AHRQ Grant Applications Submitted for Due Dates on or after May 25, 2015. December 5. Available at: https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-032.html.

NIH. 2015. NIH Databook. Available at: https://report.nih.gov/nihdatabook/. Accessed March 1, 2016.

NIH. 2016. NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information. Notice No. NOT-OD-16-149.

Normile, D. 2012. Commission spreads blame for "manmade" disaster. *Science* 337(6091): 143. July 12.

Norris, K. A. 2011. Letter to Jerome Rosenberg. September 20. Available at: old.post-gazette.com/pg/pdf/201305/norris_2011_letter_to_rosenberg.pdf. Accessed December 16, 2015.

Norris, K. A., and H. Kling. 2014. Plaintiffs' Complaint, U.S. District Court for the Western District of Pennsylvania. January 27.

Nosek, B. A., J. R. Spies, and M. Motyl. 2012. Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science* 7(6): 615-631.

Nosek, B. A., G. Alter, G. C. Banks, D. Borsboom, S. D. Bowman, S. J. Breckler, … T. Yarkoni. 2015. Promoting an open research culture. *Science* 348(6242): 1422-1425.

NPA (National Postdoctoral Association). 2013. Responsible Conduct of Research (RCR) for Post Docs. Available at: http://www.nationalpostdoc.org/publications/124-rcr. Accessed August 1.

NRC (National Research Council). 1966. Language and Machines: Computers in Translation and Linguistics. Washington, DC: The National Academies Press.

NRC. 2003. *Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences*. Washington, DC: The National Academies Press.

NRC. 2010. *NAKFI Synthetic Biology: Building a Nation's Inspiration: Interdisciplinary Research Team Summaries*. Washington, DC: The National Academies Press.

NRC. 2011. *America's Climate Choices*. Washington, DC: The National Academies Press.

NRC. 2012a. *Research Universities and the Future of America: Ten Breakthrough Actions Vital to Our Nation's Prosperity and Security*. Washington, DC: The National Academies Press.

NRC. 2012b. *Using Science as Evidence in Public Policy*. Washington, DC: The National Academies Press.

NSB (National Science Board). 1996. *Science and Engineering Indicators 1996*. Arlington, VA: National Science Foundation.

NSB. 2012. *Science and Engineering Indicators 2012*. Arlington, VA: National Science Foundation.

NSB. 2014a. *Report to the National Science Board on the National Science Foundation's Merit Review Process Fiscal Year 2014*. Available at: http://www.nsf.gov/nsb/publications/2015/nsb201514.pdf. Accessed December 13, 2015.

NSB. 2014b. *Science and Engineering Indicators 2014*. Arlington, VA: National Science Foundation.

NSB. 2014c. Reducing investigators' administrative workload for federally funded researchArlington, VA: National Science Foundation
NSB. 2016. *Science and Engineering Indicators 2016*. Arlington, VA: National Science Foundation.
NSF (National Science Foundation). 2002. Research misconduct policy. *Federal Register* 67: 11937. March 18.
NSF. 2009. NSF's Implementation of Section 7009 of the America COMPETES Act, 2009. *Federal Register* 74(160): 42126-42128.
NSF3. 2014. *Proposal and Award Policies and Procedures Guide*. Available at: http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_print.pdf.
NSF-OIG (National Science Foundation, Office of the Inspector General). 2013. Case Closeout Memorandum A13010012. Available at: https://www.nsf.gov/oig/case-closeout/A13010012.pdf.
NSF-OIG. 2015. Case Closeout Memoranda. Available at: www.nsf.gov/oig/case-closeout.jsp. Accessed December 16, 2015.
NSF-OIG. Various Years. *Semi-Annual Report to Congress*. Arlington, VA: NSF.
Nuzzo, R. 2014. Scientific method: Statistical errors. *Nature* 506(7847). Available at: http://www.nature.com/news/scientific-method-statistical-errors-1.14700.
NWC (National Whistleblowers Center). 2016. False Claims Act/Qui Tam. Available at http://www.whistleblowers.org/index.php?Itemid=64#What%20About%20State.
Nylenna, M., and S. Simonsen. 2006. Scientific misconduct: A new approach to prevention. *The Lancet* 367(9526): 1882-1884.
Obama, B. M. 2009a. Memorandum for the Heads of Executive Departments and Agencies: Scientific Integrity. March 9.
Obama, B. M. 2009b. Remarks by the President at the National Academy of Sciences Annual Meeting. April 27.
Obama, B. M. 2013. Executive Order 13642-Making Open and Machine Readable the New Default for Government Information. Available at: www.gpo.gov/fdsys/pkg/DCPD-201300318/content-detail.html. Accessed December 20, 2015.
Obokata, Haruko; Sasai, Yoshiki; Niwa, Hitoshi, et. Al. "Bidirectional developmental potential in reprogrammed cells with acquired pluripotency". Nature. **505** (7485): 676–687. PMID 24476891. doi:10.1038/nature12969.
Obokata, H., Wakayama, T., Sasai, Y., et al. (2014b) Stimulus-Triggered Fate Conversion of Somatic Cells into Pluripotency. Nature, 505, 641-647. http://dx.doi.org/10.1038/nature12968
OECD (Organisation for Economic Co-operation and Development). 2007. *Best Practices for Ensuring Scientific Integrity and Preventing Misconduct*. Paris: OECD.
OECD. 2009. *Investigating Research Misconduct Allegations in International Collaborative Research Projects: A Practical Guide*. Paris: OECD.
OECD. 2015. Research and Development Statistics. Available at: stats.oecd.org/ (accessed December 26, 2014).
Omenn, G. 2012. Evolution of Translational Omics: Lessons Learned and the Path Forward. Presentation to the committee, July 9.
Oransky, I. 2015. Lawsuit involving PubPeer unmasks commenter as pseudonymous whistleblower Clare Francis. *Retraction Watch*. April 13.
Oransky, I., and A. Marcus. 2010. Why write a blog about retractions? *Retraction Watch*. August 3.
ORI (Department of Health and Human Services, Office of Research Integrity). 1994. ORI policy on plagiarism. Office of Research Integrity Newsletter 3(1).
ORI. 2011. Historical Background. Available at: ori.hhs.gov/historical-background. Accessed January 5, 2016.
ORI. 2012-2013. Office of Research Integrity Annual Report. Washington, DC.
ORI. 2012. Office of Research Integrity Annual Report 2011. Washington, DC.
ORI. 2013. Office of Research Integrity Annual Report 2012. Washington, DC.
ORI. 2014. ORI Newsletter. November.

ORI. 2016. 5 Ways Supervisors Can Promote Research Integrity (Infographic). Available at: https://ori.hhs.gov/images/ddblock/2_Supervisor_tips.pdf.

OSC (Open Science Collaborative). 2015. Estimating the reproducibility of psychological science. *Science* 349(6251); doi: 10.1126/science.aac4716. August 28.

OSF (Open Science Framework). 2013. Reproducibility Project. Available at: http://openscienceframework.org/project/EZcUj/wiki/home. Accessed March 28, 2013.

OSTP (Office of Science and Technology Policy, Executive Office of the President). 2000. Federal policy on research misconduct. *Federal Register* 65: 76260-76264.

OSTP. 2013. Scientific Integrity. Available at: http://www.whitehouse.gov/administration/eop/ostp/library/scientificintegrity. Accessed January 10, 2013.

Pain, E. 2014. Paul Brookes: Surviving as an outed whistleblower. *Science*. March 10. Available at: http://sciencecareers.sciencemag.org/career_magazine/previous_issues/articles/2014_03_10/caredit.a1400061.

Parrish, D.M. 2005. On Identifying Research Misconduct Respondents (Letter). Science and Engineering Ethics 11, 171-172.Pascal, C. B. 1999. The history and the future of the Office of Research Integrity: Scientific misconduct and beyond. *Science and Engineering Ethics* 5(2): 183-198.

Peng, R. D. 2011. Reproducible research in computational science. *Science* 334(6060): 1226 1227.

Penslar, R. L. 1995. *Research Ethics: Cases and Materials*. Bloomington: Indiana University Press.

Peterson, R. A. 2012. Regnerus Inquiry Report. Memo. August 24. Available at: http://www.utexas.edu/opa/wordpress/news/files/Regnerus-Inquiry-Report.pdf. Accessed January 11, 2013.

Pfeffer, J., and N. Langton. 1993. The effect of wage dispersion on satisfaction, productivity, and working collaboratively: Evidence from college and university faculty. *Administrative Science Quarterly* 38(3): 382-407.

Phillips, C. J. 2012. Responsible Conduct of Research: Concerns for Postdocs. Presentation to the committee. August 14.

PhRMA (Pharmaceutical Research and Manufacturers of America). 2014. Principles on Conduct of Clinical Trials and Communication of Clinical Trial Results. Available at: http://phrma-docs.phrma.org/sites/default/files/pdf/042009_clinical_trial_principles_final_0.pdf.

Pimple, K. 2001-7. Assessing Student Learning in the Responsible Conduct of Research. Available at: http://poynter.indiana.edu/tre/kdp-assessing.pdf.

*Planned Parenthood Minn., N.D., S.D. v. Rounds,* 686 F.3d 889, 906 (8th Cir. 2012).

Plemmons, D. K., S. A. Brody, and M. W. Kalichman. 2006. Student perceptions of the effectiveness of education in the responsible conduct of research. *Science and Engineering Ethics* 12(3): 571-582.

*PNAS* (*Proceedings of the National Academy of Sciences*). 2013. PNAS Information for Authors, Revised July 2013. Available at: http://www.pnas.org/site/misc/iforc.pdf. Accessed August 1, 2013.

Popper, K. 1999. *All Life Is Problem Solving*. New York: Routledge.

Porter, A. L., and I. Rafols. 2009. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics* 81: 719-745.

Powell, S., M. Allison, and M. Kalichman. 2007. Effectiveness of a responsible conduct of research course: A preliminary study. *Science and Engineering Ethics* 13(2): 249-264..

Price, A. R. 2013. Research misconduct and its federal regulation: The origin and history of the Office of Research Integrity—with personal views by ORI's former associate director for investigative oversight. *Accountability in Research: Policies and Quality Assurance*. 20: 5-6, 291-319.

Pritchard, M. S. 2000. Service-learning and engineering ethics. *Science and Engineering Ethics* 6(3): 413-422.

Proctor, R. N. 2011. *Golden Holocaust: Origins of the Cigarette Catastrophe and the Case for Abolition*. Berkeley: University of California Press.

PSAC (President's Science Advisory Committee). 1960. *Scientific Progress, the Universities, and the Federal Government*. Washington, DC: U.S. Government Printing Office.

PubMed Commons. 2015. PubMed Commons Website. Available at: www.ncbi.nlm.nih.gov/pubmed-commons/guidelines. Accessed December 16, 2015.

RCUK (Research Councils UK). 2013. RCUK Policy and Guidelines on Governance of Good Research Conduct. February. Available at: www.rcuk.ac.uk/RCUK-prod/assets/documents/reviews/grc/RCUKPolicyandGuidelinesonGovernanceofGoodResearchPracticeFebruary2013.pdf. Accessed April 19, 2015.

Reardon, S. 2015. US vaccine researcher sentenced to prison for fraud. *Nature* 523(7559): 138-139.

Reason, J. 2000. Human error: Models and management. *BMJ* 320: 768-770.

Redman, B. K. 2013. *Research Misconduct Policy in Biomedicine: Beyond the Bad-Apply Approach*. Cambridge, MA: MIT Press.

Redman, B. K., and J. F. Merz. 2008. Scientific misconduct: Do the punishments fit the crime? *Science* 321(5890): 775.

Reiser, S. J. 1993. Overlooking ethics in the search for objectivity and misconduct in science. *Academic Medicine* 68: S84-S87.

Rekdal, O. B. 2014. Academic urban legends. *Social Studies of Science* 44(4): 638-654.

Rennie, D., and C. K. Gunsalus. 1993. Scientific misconduct: New definition, procedures, and office—perhaps a new leaf. *JAMA* 269(7): 915-917.

Rennie, D., and C. K. Gunsalus. 2001. Regulations on scientific misconduct: Lessons from the US experience, Pp. 13-31 in *Fraud and Misconduct in Biomedical Research*, 3rd Ed., S. Lock, F. Wells and M. Farthing, eds. London: BMJ Books.

Rennie, D., and C. K. Gunsalus. 2008. What is research misconduct? Pp. 29-51 in *Fraud and Misconduct in Biomedical Research,* F. Wells and M. Farthing, eds. London: Royal Society of Medicine Press.

Resnik, D. B. 2013. Plagiarism among collaborators. *Accountability in Research* 20(1): 1-4.

Resnik, D. B., and Z. Master. 2013. Policies and initiatives aimed at addressing research misconduct in high-income countries. *PLOS Medicine* 10: 3 March 26. Available at: https://doi.org/10.1371/journal.pmed.1001406.

Resnik, D. B., T. Neal, A. Raymond, and G. E. Kissling. 2015. Research misconduct definitions adopted by U.S. research institutions. *Accountability in Research: Policies and Quality Assurance* 22(1): 14-21.

Rick, S., and G. Loewenstein. 2008. Hypermotivation. *Journal of Marketing Research* 45(6): 645-648.

Rose, G. 1985. Sick individuals and sick populations. *International Journal of Epidemiology* 14: 32-38.

Rosenberg, J. L. 2011. Letter to Karen A. Norris. November 17. Available at: old.post-gazette.com/pg/pdf/201305/rosenberg_letter.pdf. Accessed December 16, 2015.

Rosenthal, R. 1979. The "file drawer problem" and tolerance for null results. *Psychological Bulletin* 86(3): 638-641.

Ross, J. S., C. P. Gross, M. M. Desai, Y. Hong, A. O. Grant, S. R. Daniels, V. C. Hachinski, R. J. Gibbons, T. J. Gardner, and H. M. Krumholz. 2006. Effect of blinded peer review on abstract acceptance. *JAMA* 295(14): 1675-1680.

Rossner, M. 2006. How to guard against image fraud. *The Scientist*. http://www.the-scientist.com/?articles.view/articleNo/23749/title/How-to-Guard-Against-Image-Fraud/.

Roth, M. 2013. University of Pittsburgh clears top researcher of alleged misconduct. *Pittsburgh Post Gazette*. http://www.post-gazette.com/news/education/2013/07/11/University-of-Pittsburgh-clears-top-researcher-of-alleged-misconduct/stories/201307110302.

Roth, M. 2014. Pitt research team sues over patent. *Pittsburgh Post-Gazette*. January 28. http://www.post-gazette.com/local/city/2014/01/27/University-of-Pittsburgh-researchers-sue-over-lung-infection-vaccine-patent/stories/201401270141.

Roth, M. 2015. Federal patent office rules against two Pitt doctors on vaccine application. *Pittsburgh Post-Gazette*. September 11. http://www.postgazette.com/news/health/2015/09/11/Federal-patent-office-rules-against-two-Pittdoctors-on-vaccine-application/stories/201509110323.

Roth, M., and B. Schackner. 2013. Pitt scientist Jay Kolls faces another investigation. *Pittsburgh Post-Gazette*. http://www.post-gazette.com/news/education/2013/05/26/Pitt-scientist-Jay-Kolls-faces-another-investigation/stories/201305260225.

Royal Society Science Policy Center. 2012. *Science as an Open Enterprise*. Final Report. London: Royal Society. Available at: https://royalsociety.org/~/media/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf.

Ruedy, N. E., F. Gino, C. Moore, and M. E. Schweitzer. 2013. The cheater's high: The unexpected affective benefits of unethical behavior. *Journal of Personality and Social Psychology* 105(4): 531-548.

Sampson, R. J., and J. H. Laub. 1992. *Crime in the Making*. Cambridge, MA: Harvard University Press.

Sauermann, H., and M. Roach. 2012. Science PhD career preferences: Levels, changes, and advisor encouragement. *PLOS ONE* 7(5): e36307.

Scargle, J. 2000. Publication bias: The "file-drawer" problem in scientific inference. *Journal of Scientific Exploration* 14(1): 91-106.

Schachman, H. K. 1993. What is misconduct in science. *Science* 261(5118):148-149.

Schillebeeckx, M., B. Maricque, and C. Lewis. 2013. The missing piece to changing the university culture. *Nature Biotechnology* 31: 938-941.

Schmaling, K. B., and A. W. Blume. 2009. Ethics instruction increases graduate students' responsible conduct of research knowledge but not moral reasoning. *Accountability in Research: Policies & Quality Assurance* 16(5): 268-283.

Schneider, S. L., K. K. Ness, S. Rockwell, and K. Shaver. 2012. *2012 Faculty Workload Survey: Research Report*. Washington, DC: Federal Demonstration Partnership. Available at: sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga_087667.pdf. Accessed December 26, 2014.

Schwartz, L. M., S. Woloshin, A. Andrews, and T. A. Stukel. 2011. Influence of medical journal press releases on the quality of associated newspaper coverage: Retrospective cohort study. *BMJ* 344: d8164.

Science. 2011. Dealing with Data (Special Issue). 331:6018. February 11, 2011.

Science. 2012. Newsmakers. 335(6066): 270-271. January 20.

Seife, C. 2015. Research misconduct identified by the US Food and Drug Administration: Out of sight, out of mind, out of the peer-reviewed literature. *JAMAJAMA Internal Medicine* 175(4): 567-577.

Seiler, S. N., B. J. Brummel, K. L. Anderson, K. J. Kim, S. W. Wee, C. K. Gunsalus, and M. C. Loui. 2011. Outcomes assessment of role-play scenarios for teaching responsible conduct of research. *Accountability in Research* 18: 217-246.

Servick, K. 2015. Michigan judge asks PubPeer to turn over anonymous user information. *Science*. http://news.sciencemag.org/people-events/2015/03/michigan-judge-asks-pubpeer-turn-over-anonymous-user-information.

SfN (Society for Neuroscience). 2010. Guidelines for Responsible Conduct Regarding Scientific Communication. Available at: http://www.sfn.org/Member-Center/Professional-Conduct/Guidelines-for-Responsible-Conduct-Regarding-Scientific-Communication. Accessed July 23, 2013.

Shea, C. 2012. The data vigilante. *The Atlantic*. December.

Sieber, J. E. 2005. Misconceptions and realities about teaching online. *Science and Engineering Ethics* 11(3): 329-340.

Siegrist, J. 2001. A theory of occupational stress. In *Stress in the Workplace: Past, Present and Future*, J. Dunham, ed.). London: Whurr.

Siegrist, J., N. Wege, F. Pühlhofer, and M. Wahrendorf. 2008. A short generic measure of work stress in the era of globalization: Effort-reward imbalance. *International Archives of Occupational and Environmental Health* 82(8): 1005-1013.

Simonsohn, U. 2013. Just post it: The lesson from two cases of fabricated data detected by statistics alone. *Psychological Science* 24: 1875-1888.

*Singapore Statement on Research Integrity*. 2010. 2nd World Conference on Research Integrity. Available at http://www.singaporestatement.org/index.html. Accessed July 26, 2012.

Smith, E., and B. Williams-Jones. 2012. Authorship and responsibility in health sciences research: A review of procedures for fairly allocating authorship in multi-author studies. *Science and Engineering Ethics* 18(2): 199-212.

Smith, K., D. Wueste, and J. Frugoli. 2007. Using "ethics labs" to set a framework for ethical discussion in an undergraduate science course. *Biochemistry and Molecular Biology Education* 35: 332-336.

Smith, R. 2000. What is research misconduct? Pp. 7-11 in *The COPE Report 2000: Annual Report of the Committee on Publication Ethics*, C. White, ed. London: BMJ Books.

Smith, R. 2006. Research misconduct: The poisoning of the well. *Journal of the Royal Society of Medicine* 99(5): 232-237.

Smith, R. 2013. Should scientific fraud be a criminal offense? *BMJ Blog*. December 9. Available at: blogs.bmj.com/bmj/2013/12/09/richard-smith-should-scientific-fraud-be-a-criminal-offence/. Accessed December 17, 2015.

Sovacool, B. 2008. Exploring scientific misconduct: Isolated individuals, impure institutions, or an inevitable idiom of modern science? *Journal of Bioethical Inquiry* 5(4): 271-282.

Sox, H. C., and D. Rennie. 2006. Research misconduct, retraction, and cleansing the medical literature: Lessons from the Poehlman case. *Annals of Internal Medicine* 144(8): 609-613.

SPSP (Society for Personality and Social Psychology). 2012. *Report of the SPSP Task Force on Responsible Conduct in Research*. June 10.

SRCR (Secretariat on Responsible Conduct of Research). 2016. Tri-Agency Framework: Responsible Conduct of Research. Available at: http://www.rcr.ethics.gc.ca/eng/policy-politique/framework-cadre/.

Steen, R. G. 2011. Retractions in the scientific literature: Is the incidence of research fraud increasing? *Journal of Medical Ethics* 37(4): 249-253.

Steen, R. G., A. Casadevall, and F. C. Fang. 2013. Why has the number of scientific retractions increased? Plosone8.e68397.10.1371/journal.pone.0068397.

Steinbrook, R., and R. F. Redberg. 2015. Reporting research misconduct in the medical literature. *JAMA Internal Medicine* 175(4): 492-493.

Steneck, N. H. 1994. Research universities and scientific misconduct—history, policies, and the future. *Journal of Higher Education* 65(3): 310-330.

Steneck, N. H. 2004. *ORI Introduction to the Responsible Conduct of Research*. Washington, DC: U.S. Government Printing Office.

Steneck, N. H. 2006. Fostering integrity in research: Definitions, current knowledge, and future directions. *Science and Engineering Ethics* 12(1): 53-74.

Steneck, N. H. 2007. *Introduction to the Responsible Conduct of Research*. Office of Research Integrity, U.S. Department of Health and Human Services. Washington, DC: U.S. Government Printing Office.

Stephan, P. 2012a. *How Economics Shapes Science*. Cambridge, MA: Harvard University Press.

Stephan, P. 2012b. Research efficiency: Perverse incentives. *Nature* 484: 29-31. April.

Sterman, J. D. 1994. Learning in and about complex systems. *System Dynamics Review* 10(2-3): 291-330.

Stern, S., and T. Lemmens. 2011. Legal remedies for medical ghostwriting: Imposing fraud liability on guest authors of ghostwritten articles. *PLOS Medicine* 8(8): e1001070.

Stodden, V. 2009. Enabling reproducible research: Licensing scientific innovation. *International Journal of Communications Law & Policy* 13(Winter): 22-46.

Stodden, V. 2010. Open science: Policy implications for the evolving phenomenon of user-led scientific innovation. *Journal of Science Communication* 9(1): 1-8.

Stroebe, W., T. Postmes, and R. Spears. 2012. Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science* 7(6): 670-688.

Strohmetz, D. B. 1992. The use of role-play in teaching research ethics: A validation study. *Teaching of Psychology* 19(2): 106.

TAFC (Tenure and Academic Freedom Committee, University of Pittsburgh). 2013. Letter to Karen A. Norris. March 4. Available at: old.post-gazette.com/pg/pdf/201305/tafc_letter_to_norris.pdf. Accessed December 16, 2015.

Taichman, D. B., J. Backus, C. Baethge, H. Bauchner, P. W. de Leeuw, J. M. Drazen, J. Fletcher, F. A. Frizelle, T. Groves, A. Haileamlak, A. James, C. Laine, L. Peiperl, A. Pinborg, P. Sahni, and S. Wu. 2016. Sharing clinical trial data: A proposal from the International Committee of Medical Journal Editors. *JAMA* 315(5): 467-468.

Tatsioni, A., N. G. Bonitsis, and J. P. A. Ioannidis. 2007. Persistence of contradicted claims in the literature. *JAMA* 298(21): 2517-2526.

Taylor, B., E. Miller, C. P. Farrington, M. C. Petropoulos, I. Favot-Mayaud, J. Li, and P. A. Waight. 1999. Autism and measles, mumps, and rubella vaccine: No epidemiologic evidence for a causal association. *The Lancet* 353: 2026-2029.

Teitelbaum, M. S. 2008. Structural disequilibria in biomedical research. *Science* 321(5859): 644-645.

TENK (National Advisory Board on Research Ethics). 2002. *Good Scientific Practice and Procedures for Handling Misconduct and Fraud in Science*. Finland.

Thaler, R. H., and C. R. Sunstein. 2008. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Thiese, M. S., Z. C. Arnold, and S. D. Walker. 2015. The misuse and abuse of statistics in biomedical research. *Biochemia Medica (Zagreb)* 25(1): 5-11.

Titus, S. L., and J. M. Ballou. 2014. Ensuring PhD development of responsible conduct of research behaviors: Who's responsible? *Science and Engineering Ethics*. 20(1): 221-235.

Titus, S. L., J. A. Wells, and L. J. Rhoades. 2008. Repairing research integrity. *Nature* 453: 980-982.

Triggle, N. 2010. MMR doctor struck from register. *BBC News*. May 24. Available at: http://news.bbc.co.uk/2/hi/health/8695267.stm. Accessed November 12, 2016.

Tyersky, A., and D. Kahneman. 1974. Judgement under uncertainty: Heuristics and biases. *Science* 185(4157): 1124-1131.

Tyler, T. R., and S. L. Blader. 2003. The Group Engagement Model: Procedural justice, social identity, and cooperative behavior. *Personality and Social Psychology Review* 7(4): 349-361.

UK GMC (United Kingdom General Medical Council). 2010. Fitness to Practice Panel Hearing. 28 January.

UKRIO (United Kingdom Research Integrity Office). 2009. *Code of Practice for Research: Promoting Good Practice and Preventing Misconduct*. London: UKRIO.

UKRIO. 2017. About Us. Available at: http://ukrio.org/about-us/.

University of Illinois. 2009. Policy and Procedures on Integrity in Research and Publication.

University of Kansas. 2013. Public Censure Notice. December 11. Available at: news.ku.edu/2013/12/11/public-censure. Accessed December 17, 2015.

UUK (Universities UK). 2012. The Concordat to Support Research Integrity. Available at: http://www.universitiesuk.ac.uk/policy-and-analysis/reports/Documents/2012/the-concordat-to-support-research-integrity.pdf.

Van Noorden, R. 2011. The trouble with retractions. *Nature* 478: 26-28.

Vaughan, D. 1999. The dark side of organizations: Mistake, misconduct, and disaster. *Annual Review of Sociology* 25: 271-305.

Vogel, G., and D. Normile. 2014. Exclusive: Nature reviewers not persuaded by initial STAP stem cell papers. *Science*. September 11. Available at: http://news.sciencemag.org/asiapacific/2014/09/exclusive-nature-reviewers-not-persuaded-initial-stap-stem-cell-papers.

Wager, E. 2014. Defining and responding to plagiarism. *Learned Publishing* 27: 33-42.

Wager, E. 2015a. How should journal editors respond to cases of suspected misconduct? *Journal of Microbiology & Biology Education* 15(2): 146-150.

Wager, E. 2015b. Why are retractions so difficult? *Science Editing* 2(1): 32-34.

Wager E., and S. Kleinert. 2012. *Cooperation Between Research Institutions and Journals on Research Integrity Cases: Guidance from the Committee on Publication Ethics (COPE)*. www.publicationethics.org.

WAME (World Association of Medical Editors). 2007. Authorship. Available at: www.wame.org/about/policy-statements#Authorship, Accessed June 16, 2016.

Wansink, B., and K. van Ittersum. 2006. The visual illusions of food: Why plates, bowls, and spoons can bias consumption volume. *FASEB Journal* 20: A618.

Washington University. 2009. Washington University Policy for Authorship on Scientific and Scholarly Publications. Available at: http://research.wustl.edu/PoliciesGuidelines/Pages/AuthorshipPolicy.aspx.

Wasserstein, R. L., and N. A. Lazar. 2016. The ASA's statement on *p*-values: context, process, and purpose. *American Statistician* 70(2): 129-133.

Weed, D. L. 1998. Preventing scientific misconduct. *American Journal of Public Health* 88(1): 125-129.

Wells, J. A., C. R. Thrush, B. C. Martinson, T. A. May, M. Stickler, E. C. Callahan, and K. L. Klomparens. 2014. Survey of organizational research climates in three research intensive, doctoral granting universities. *Journal of Empirical Research on Human Research Ethics* 9(5): 72-88.

Wester, K. L., J. T. Willse, and M. S. Davis. 2008. Responsible conduct of research measure: Initial development and pilot study. *Accountability in Research: Policies and Quality Assurance* 15(2): 87-104.

White, C. 2005. Suspected research fraud: Difficulties of getting at the truth. *BMJ* 331(7511): 281-288.

White, P. F., H. Kehlet, and S. Liu. 2009. Perioperative analgesia: What do we still know? *Anesthesia & Analgesia* 108(5): 1364-1367.

Wickenden, W. E. 1949. *A Professional Guide for Young Engineers*. Engineers' Council for Professional Development. Available at: http://www.scribd.com/doc/58348312/A-Professional-Guide-for-Young-Engineers-William-E-Wickenden#scribd. Accessed March 22, 2015.

Wilhite, A. W., and E. A. Fong. 2012. Coercive citation in academic publishing. *Science* 335(6068): 542-543.

Williams, E. 2012. Industry Perspectives on Research Integrity—Part One. Presentation to the Committee. August 14.

Wislar, J. E., A. Flanagin, and P. B. Fontanarosa. 2011. Honorary and ghost authorship in high impact biomedical journals: A cross sectional survey. *BMJ* 343: d6128.

Wolfe-Simon, F., J. S. Blum, T. R. Kulp, G. W. Gordon, S. E. Hoeft, J. Pett-Ridge, J. F. Stolz, S. M. Webb, P. K. Weber, P. C. W. Davies, A. D. Anbar, and R. S. Oremland. 2011. A bacterium that can grow by using arsenic instead of phosphorus. *Science* 332(6034): 1163-1166.

Woolston, C. 2002. When a mentor becomes a thief. *Science Magazine* (Online). Available at: http://www.sciencemag.org/careers/2002/05/when-mentor-becomes-thief.

Wright, D. E., S. L. Titus, and J. B. Cornelison. 2008. Mentoring and research misconduct: An analysis of research mentoring in closed ORI cases. *Science and Engineering Ethics* 14(3): 323-336. September.

Wuchty, S., B. F. Jones, and B. Uzzi. 2007. The increasing dominance of teams in production of knowledge. *Science* 316: 1036-1039.

Yank, V., and D. Rennie. 1999. Disclosure of researcher contributions: A study of original articles in *The Lancet*. *Annals of Internal Medicine* 130(8): 661-670.

Yap, A. J., A. S. Wazlawek, B. J. Lucas, A. J. C. Cuddy, and D. R. Carney. 2013. The ergonomics of dishonesty: The effect of incidental posture on stealing, cheating, and traffic violations. *Psychological Science* 24(11): 2281-2289.