Ethical Reasoning for a Data-Centered World

By

Rochelle E. Tractenberg, PhD, MPH, PhD, PStat®, FASA, FAAAS

Ethical Reasoning for a Data-Centered World

By Rochelle E. Tractenberg, PhD, MPH, PhD, PStat®, FASA, FAAAS

This book first published 2022

Ethics International Press Ltd, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2022 by Rochelle E. Tractenberg

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical photocopying, recording or otherwise, without the prior permission of the copyright owner.

Print Book ISBN: 978-1-80441-078-3

eBook ISBN: 978-1-80441-079-0

Dedication

This book is dedicated to my fellow practitioners, and to instructors who have an interest in integrating ethical reasoning, and ethics, into their courses on statistics, computing, and data science. I know that students and practitioners who work with data directly or indirectly may share this interest at different times in their careers and training; I hope the book serves instructors, learners, and practitioners equally well in their commitment to ethical engagement with data.

Acknowledgements

This is my first book, the first draft of which was completed during a sabbatical from Georgetown University (2019). My ethical reasoning mentor, Father Kevin FitzGerald SJ, PhD, PhD, was instrumental in forging my commitment to ethical reasoning for teaching, learning, and enabling others to develop an ethical professional identity. Father Kevin's friendship, optimism, and encouragement have shone some of the brightest lights on my two decades (so far!) at Georgetown.

While the book is focused on ethical reasoning, and can be used to build its requisite knowledge, skills, and abilities generally, it is structured around statistics, computing, and data science. While all errors are definitely my own, the work benefitted greatly from frequent, fun, and challenging discussions with Donna LaLonde of the American Statistical Association about the obstacles that instructors in statistics and data science might face when they seek to integrate ethical reasoning and ethical content into their courses. (The conversations about dogs and running were also very helpful!) Donna's relentlessly concrete suggestions and honest feedback have made her a great friend as well as an awesome sounding board. I am lucky to have the benefit of her experience, expertise, and encouragement.

The Committee on Professional Ethics (COPE) of the American Statistical Association, which I had the privilege of vice-chairing (2014-2016) and chairing (2017-2019), was formative in my commitment to writing this book in order to help all instructors to utilize the ASA Ethical Guidelines for Statistical Practice. I am grateful to Howard Hogan, fellow Camarillo resident, for his faith in me and encouragement to fully engage with the Committee's charter.

Stewardship of the profession and discipline of science has been a feature of my professional identity for decades. I owe a debt of gratitude to the Carnegie

Initiative on the Doctorate and Chris Golde in particular for bringing the construct of disciplinary stewardship into being. Dr. Golde, and Dr. Chris Rios, facilitated my thinking about expanding stewardship beyond doctoral training and to the profession. I hope this book can serve to engage readers with disciplinary and professional stewardship as they learn to reason ethically in the modern, data-centered, world.

Table of Contents

List of Cases

Abbreviations:

DEFW is the Data Ethics Framework

DSEC is the Data Science Ethics Checklist

ASA is the American Statistical Association

ACM is the Association of Computing Machinery

Page	Case Title	Relevant Principles
288	**Case:** You are directed to design a system to scrape data from a specific source (e.g., Facebook), and are provided with specific design features of the source to ensure every data type can be scraped from every user.	**ASA** A, D, E, F, H **ACM**: 1, 2
298	**Case:** You build a data scrape algorithm with a built-in opt-in feature, that will pop up and ask the user to opt-in to the data scraping (i.e., give consent) every time the algorithm changes, to scrape/collect more, or different data. That feature is removed by someone.	**ASA**: A, D, E, F, H **ACM**: 1, 2
310	**Case:** Piles of data start arriving from "the company data scraper" for you to analyze. You begin to run existing analysis programs, but discover that no information is available to indicate whether consent was given for any of the data. You suspect, and then potentially identify, an error in the analysis code that removed that information.	**ASA**: A, B, D, E, F, H **ACM**: 1, 2, 3

Page	Case Title	Relevant Principles
321	**Case:** Results suggest that some people who are the source of data your organization wants to use (e.g., Facebook) are more susceptible to messaging (e.g., advertisements and fake news items) than others. You interpret this as signalling a need for caution/care in what your system does next with this data, and you include this in all your communication – in order to limit bias, ensure that no stakeholders are misled, and support valid conclusions resulting from your statistical, computing, and data science practice. Instead, you find reports of your work interpret the results without any caveats, and remove any suggestions that caution or sensitivity analyses may be needed.	**ASA**: A, B, C, D,E, F, G, H **ACM**: 2
331	**Case:** You are told **not** to document your work. When you do (because that's what the practice standards say the ethical practitioner does), your boss/supervisor returns it to you with the direction, "fix this".	**ASA**: A, B, C, E, F, G, H **ACM**: 1, 2, 3, 4
340	**Case:** You submit your complete and correct report of your scraping algorithm – including identification of the removal of your built in, opt-in consent to contribute data; the lack of consent accompanying data to be analyzed; and the lack of your recommendations in interpretations for limiting bias. You later discover that none of this documentation was included in the final report, but the final report is shared with stakeholders as if it is complete and correct.	**ASA**: A, B, C, D, E, F, G, H **ACM**: 1, 2, 3, 4
355	**Case:** Leadership informs your team that they bought an algorithm that you will be using to scrape data. But first, they want you to take off all the consent pop-ups, because "that ruins the user experience" and "adds personal data we will only need to strip off to preserve confidentiality".	**ASA**: A, B, C, E, G, H **ACM**: 1, 2, 3, 4

Introduction

The subject of the book is *ethical reasoning*: how to do it, why learn to do it, and how to teach and learn to do it, and document that it has been learned (and improved). Ethical reasoning is its own set of knowledge, skills, and abilities (KSAs (Santa Clara University (no date); Tractenberg & FitzGerald, 2012; Tractenberg et al. 2017)). These KSAs are learnable and improvable, and can be deployed to ensure ethical practice (when there is no/before there is an ethical problem about which a decision has to be made) as well as when a decision about what to do (ethically) is required. Thus, learning to reason ethically – rather than "learning the Ethical Guidelines and/or Code of Ethics" – will promote "…the skill, good judgment, and polite behavior that is expected from a person who is trained to do a job well" more generally, and more universally (see Rios et al. 2019).

The academic level at which the book is targeted is anyone who is preparing to engage in statistics or data science in the course of their work – whether that is their main task or simply a toolset they will utilize. This book introduces and discusses ethical reasoning as a learnable, improvable skill set that an individual can learn themselves, and document, *and* which an instructor can teach and assess (and develop themselves). The context of the book is "the data-centered world", wherein "quantitative practice" is any person's applications of statistics, data science, or some combination of these. Importantly, an individual who does one "simple analysis" is engaging in quantitative practice, although not to the extent that someone with a job title "statistician" does; as you will see in later chapters, practice standards are relevant for both of these individuals – making this book, and its application of those practice standards across tasks, relevant right across quantitative practice – however extensive or limited it may be.

The material and examples should be accessible to advanced undergraduate or graduate students, as well as practitioners. This book presents the current (2022) ASA Ethical Guidelines for Statistical Practice[1], which are emphatically relevant for <u>any</u> person analyzing data, whether they are a researcher/scientist or not; and whether or not the practitioner self-identifies as a statistician. Also presented are the current (2018, draft 3) Code of Ethics of the Association of

---

[1] https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx

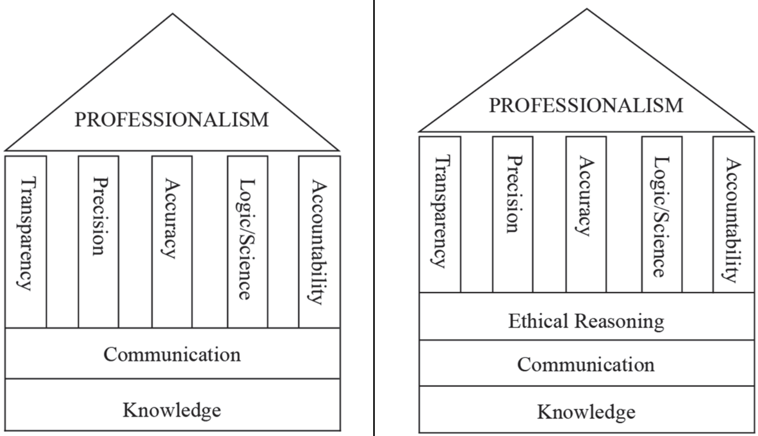
Computing Machinery[2]. The book could be used for a stand-alone course, as an adjunct text for a consulting course, or for use alongside an internship and/or capstone course (for advanced undergraduates or graduate students) in any discipline or program where data are utilized.

How Ethical Reasoning Figures in Professional Identity Development

Figures 1A and 1B[3] show two alternative models of professionalism specific to quantitative sciences. The figures are adapted by the author from a model for professionalism in the health professions (Figure B, Arnold & Stern, 2006). In the figure below, the five key elements of competent work by statisticians (identified by Rosnow & Rosenthal, 2011) are shown as pillars supporting "professionalism": transparency, precision, accuracy, accountability, and logical and scientific justification <for work>. There is no mention of ethical knowledge or reasoning in any discussion of professionalism among those who use statistics and data science –whether this is a primary or secondary/ancillary part of their work. The American Statistical Association (ASA) Ethical Guidelines for Statistical Practice (or GLs) and the Association for Computing Machinery (ACM) Code of Ethics (or CE) implicitly blend these two constructs – ethics and professionalism. Both of these professional practice standards are also explicit in their applicability to any person, irrespective of job title or educational background, that will utilize the tools, methods, and constructs of these professions.

[2] https://www.acm.org/about-acm/code-of-ethics

[3] The figures and descriptive paragraphs are excerpted from Tractenberg RE. (2013). Ethical Reasoning for Quantitative Scientists: A Mastery Rubric for Developmental Trajectories, Professional Identity, and Portfolios that Document Both. Proceedings of the 2013 Joint Statistical Meetings, Montreal, Quebec, Canada. Pp. 3959-3973.



A. A model that excludes ethical reasoning in the development of professionalism. Knowledge and communication skills are the foundation for transparency, precision, accuracy, accountability, and logical and scientific justification of competent statistical work, which in turn are the pillars supporting professionalism.

B. A model of how ethical reasoning can be used to build professionalism. Knowledge and communication skills, *when combined with ethical reasoning*, are the foundation of transparency, precision, accuracy, accountability, and logical and scientific justification of competent statistical work, the pillars that support professionalism.

Figures 0.1A-0.1B: Two alternative models of professionalism specific to quantitative sciences; adapted from Figure B, Arnold & Stern (2006).

The models shown in Figures 0.1A and 0.1B support the conceptualization of ethical reasoning as integral for both effective work (e.g., Rosnow & Rosenthal, 2011) and the development of a sense of professionalism (e.g., Stern, 2006). For mathematics education, Ferrini-Mundy (2008) mentions "ethics" as an element of the core knowledge that PhD students in mathematics need, but does not discuss whether this can or should be integrated into this doctoral training as integral to the formation of the mathematician's professional identity. For undergraduate Data Science education, the National Academies reported that "…students will need exposure to material from multiple disciplines…and they will need training in …ethical problem solving." This book is intended to provide such training, to prepare those who will use math, statistics, and data science to do so ethically in a data-centered world.

Organization of the book

The book is structured around engaging the reader in increasingly sophisticated engagement with ethical reasoning (ER) and the professional practice standards. First, familiarity is built with these standards and how they can create a decision-making framework that a quantitative practitioner can always utilize (Section 1). In Section 2, this engagement deepens to allow the reader to understand and gain practice in how the standards – current and future revisions - can be used to plan quantitative work, roughly sorted into six tasks, plus teamwork (for a total of seven tasks). In the final section, the deepest level of engagement involves the reader learning how to formulate and justify decisions about responses that might be warranted to ethical challenges that can arise (and have arisen!) in the data-centered world.

Section 1. Background and introduction

This Section provides the background information as an introduction to the book and its structure. Part of this information is Ethical Reasoning (ER), which comprises six elements of knowledge, skill, and ability (KSAs). Part of the background are the ethical practice standards of statistical and data science practice from the American Statistical Association (2022), and the Association of Computing Machinery (2018). Other support for ER includes the Stakeholder Analysis (SHA; Tractenberg, 2019-d), plus other key topics of interest for ethical engagement with data. Chapters 1.3-1.6 are all examples of "prerequisite knowledge", which is the first KSA of Ethical Reasoning (ER KSA 1), and its central importance in ethical reasoning. KSA 2 relates to choosing an ethical framework from which decisions can be made. Even before any ethical challenge or problem arises, a better understanding of why the GLs/CE exist can be gained by exploring two ethical frameworks that can be useful in decision making: "virtue ethics" and "utilitarian ethics". This prerequisite knowledge is summarized in Chapter 1.7, with additional information shared in Chapter 1.8. In Chapter 1.9, the reader moves on to ER KSAs 3, 4 (Chapter 1.10), with 5 and 6 presented together in Chapter 1.11.

Chapter 1.1. Introduction and Background

Introducing frameworks for ethical practice in statistics and data science

Introduce Bloom's taxonomy of cognitive behaviors (Bloom et al. 1956), explain its relevance for the book. *This Section is focused on Bloom's 1-4*

Chapter 1.2. Ethical Reasoning

Ethical Reasoning (ER) comprises six elements of knowledge, skill, and ability (KSAs). This chapter introduces ER and its importance, contrasting it with other "methods" for identifying, making decisions about, and justifying decisions about ethical challenges that arise whenever data are involved. Engagement with data goes beyond just data analysis –whether the individual analyzes the data formally themselves (e.g., as a statistician) or whether algorithms are used or created to analyze the data using software or other technology (e.g., as a data scientist).

Chapter 1.3. ASA Ethical Guidelines

This chapter presents the Guidelines (GLs) for Ethical Practice from the American Statistical Association (ASA, 2022).

Chapter 1.4. ACM Code of Ethics

This chapter presents the Association for Computing Machinery (ACM, 2018) Code of Ethics (CE).

Chapter 1.5. Prerequisite knowledge common to ACM CE and ASA GLs

This chapter presents the alignment – or concordance - of ASA and ACM Guidelines/Code, summarizing prerequisite knowledge that is common to both ACM CE and ASA GLs, and the "virtue ethics" framework for decision making. While not proposing that the virtue ethics model is better or worse than any other, understanding what this framework provides in terms of decision-making support is important for a full understanding of ethical reasoning.

Chapter 1.6. Stakeholder Analysis and Utilitarian Decision-making Framework

The Stakeholder Analysis and Utilitarian framework for decision making are presented. The utilitarian perspective may be a familiar one to readers, but how it can interact with professional practice standards will not be (because such standards do not typically integrate decision making frameworks). A key challenge in implementing the Utilitarian decision-making framework can be in the recognition and estimations of "harms" and "goods" or benefits, and the stakeholder analysis can help us to do this.

Chapter 1.7. Summarising KSAs 1 & 2

This chapter recaps the prerequisite knowledge and the KSAs of ER, focusing on prerequisite knowledge and decision-making frameworks, KSAs 1-2.

Chapter 1.8. Aligning prior training to promote ethical qualitative practice

This chapter discusses "other" ethics training readers in the US may have had, from the perspectives (and according to the priorities) of the US National Institutes of Health and National Science Foundations. These topics represent ancillary prerequisite knowledge, as well as definitions of human subjects, "research", and the concept of informed consent represent norms that are essential to understanding basic human rights (particularly of autonomy) and how these are addressed in the ethical practice standards of the ASA and ACM.

Chapter 1.9. Identify or recognize the ethical issue KSA 3

Returning to the ER KSAs, Chapter 1.9 discusses KSA 3, identifying or recognizing the ethical issue in any case or situation. This is one of the most difficult parts of the ER process, and is crucially dependent on a firm understanding of ethical practice standards like the GLs and CE, and also utilizes the Stakeholder Analysis (SHA).

Chapter 1.10. Identify alternative actions KSA 4

This chapter focuses on using the ethical practice standards to accomplish KSA 4, identify alternative actions (on the ethical issue), is only slightly less difficult that KSA 3. Two examples demonstrate how "do nothing" or "ignore the situation" are contrary to both the ACM and ASA standards, but offer a general approach to identifying plausible options for responding to an ethical issue (KSA 3) that are consistent with the standards.

Chapter 1.11. Make and justify a decision and reflect on it KSAs 5 & 6

The final ER KSAs, make and justify decision (KSA 5), and reflect on that decision (KSA 6) are discussed, in terms of how they represent (concretely) ethical practice, and also how they have the potential to influence "norms" in the workplace.

Section 2. Establishing familiarity with ASA and ACM principles/elements as they relate to the tasks (anticipating what problems may arise)

Introduction to Section 2: This section focuses the reader's development of cognitive capabilities on Bloom's Taxonomy levels 3-5, while reinforcing Bloom's 1-2 level familiarity with the practice standards. Emphasis is on understanding the relevance of GL and CE content for each of the tasks involved in work with data, leaning towards synthesis (of your experience with your new knowledge; or of different/diverse types of knowledge). Section 2 directs attention to how the SHA, GL, and CE can all provide guidance for your thinking/planning/workflows, as well as for each of the typical tasks in statistics and data science. Structure of the section is to introduce a SHA for each task. This raises the reader's Bloom's level (because SHA requires *analysis* (Bloom's 3) and *predictions* (Bloom's 4)) while reinforcing the reader's familiarity with the core content (Bloom's 1-2).

In each of the Section 2 chapters (2.2-2.9), both the ASA and ACM guidance on each task is identified. A SHA is completed, outlining harms and benefits that accrue when the task is completed with/without compliance with the ethical practice standards. Discussion questions are included in each chapter.

Section 3. Ethical reasoning using ASA and ACM principles/ elements: vignettes

Chapter 3.1. Introduction to Section 3.

This section focuses the reader's development of Bloom's 4-6 thinking while reinforcing Bloom's 1-5 level engagement with the practice standards. In Section 3, there are ethical challenges that need identifying (ER KSA 3) and these come from an evaluation (Bloom's 5) of a brief vignette describing workplace events.

In this section the emphasis is on responding to the ethical challenge (i.e., using ER KSAs 3-6), requiring synthesis of prerequisite knowledge (GLs/CE) and understanding how the SHA drives the decisions as well as the justifications for how to respond to these workplace situations. The vignettes describe actual events relating to each of the typical tasks in statistics and data science (as introduced in Section 2). The vignettes are analyzed following the ER KSAs, including reflection that combines what a typical practitioner might consider about the case, and also thoughts about what the decision means from an instructional perspective. In this section, all KSAs of ethical reasoning are re-introduced and utilized/practiced.

Chapter 3.2. Planning/Designing

You are directed to design a system to scrape data from a specific source (e.g., Facebook), and are provided with specific design features of the source to ensure every data type can be scraped from every user.

Chapter 3.3. Data collection/munging/wrangling

You build a data scrape algorithm with a built-in opt-in feature, that will pop up and ask the user to opt-in to the data scraping (i.e., give consent) every time the algorithm changes, to scrape/collect more, or different data. That feature is removed.

Chapter 3.4. Analysis (perform or program to perform)

Piles of data start arriving from "the company data scraper" for you to analyze. No information is available that indicates whether consent was given for any of the data. You suspect, and then potentially identify an error in the code that removed that information.

Chapter 3.5. Interpretation

Results suggest that some people using the source (e.g., Facebook) are more susceptible to messaging (e.g., advertisements and fake news items) than others. You interpret this as signaling a need for caution/care in what your system does next – in order to limit bias, and support valid conclusions. Instead, you find reports of your work interpret the results without any caveats, and have removed any of your suggestions that sensitivity analyses may be needed.

Chapter 3.6. Documenting your work

You are told not to document your work. When you do (because that's what the practice standards say the ethical practitioner does), your boss returns it to you with the direction, "fix this".

Chapter 3.7. Reporting your results/communication

You submit your complete and correct report of your scraping algorithm – including identification of the removal of your built in, opt-in consent to contribute data; the lack of consent accompanying data to be analyzed; and the lack of your recommendations in interpretations for limiting bias. You later discover that none of this documentation was included in the final report, but the final report is shared with stakeholders as if it is complete and correct.

Chapter 3.8. Engaging in team science/teamwork

Leadership informs your team that they bought an algorithm that you will be using to scrape data. But first, they want you to take off all the consent pop-ups, because "that ruins the user experience" and "adds personal data we will only need to strip off to preserve confidentiality".

Chapter 3.9. Embracing your inner ethical practitioner

This chapter invites the reader to return to each of the chapters and vignettes in Section 3. Each case analysis that is provided is revisited through the lens of ***role playing***. Specifically, readers are invited to practice – in actual role-playing dyads, or in writing, or both – the delivery of their ethically-reasoned decision in the workplace. Readers may feel more confident in engaging in conversations with others with a full, written analysis of what the issue is/issues are, what alternatives were considered, what decision is recommended/was taken and why, and some reflection on the relevance of the decision. This engagement is a specific part of ethical practice according to the

ACM and ASA, and is encouraged by the National Academy of Engineering (2013) and National Academies of Science, Engineering, and Medicine (NASEM, 2017). Initiating, and/or engaging in, that kind of conversation requires at least some practice! Readers should role play both sides of the conversation:

- describe your analysis of the case and your decision;
- "receive" a case analysis, and collaboratively determine the best way to ensure such situations do not recur;
- "receive" a case analysis and try to dissuade the analyzer from making – or publicizing – that decision;
- respond to someone who does not support your analysis and your decision.

Chapter 3.10. Summary of Section 3 and the book: career spanning engagement in professional and ethical practice

Reconsider what you have learned, and go on to refine what you have learned.

In each chapter of Section 3, we analyzed one vignette per task and readers were invited to analyze that case analysis. The objective of the Section was to teach and give practice in the full range of KSAs that are required for ethical reasoning – and not to make sure every conceivable situation was explored. The reader is encouraged to continue with self-directed learning, using the same KSAs on new problems and in particular, reflecting on the utility of each analysis to improve the chances of an *ethical* data-centered world. To support this self-direction, the reader is invited to consider, and discuss, how the ASA Guidelines and ACM Code of Ethics promote professionalism (in you) or the profession of statistics and data science (more generally). For example, consideration of the following:

- Explain whether/how the application of the Guidelines in any given case encourages ethical conduct in research (or practice, as appropriate) more generally.

- Do the ASA Guidelines and/or ACM Code of Ethics promote *professionalism*? How/how not?

Section 1. Background and introduction

Chapter 1.1
Introducing frameworks for ethical practice in statistics and data science[4]

1. Ethical practice = Frameworks + Reasoning

Data science is a new discipline, but it arises from two disciplines with long-standing commitments to ethical practice: computing and statistics. Ethical guidelines have been developed over several decades to support ethical professional practice with – as well as the application of – tools, techniques, and methods from both statistics and computing (Tractenberg et al. 2015; see also Tractenberg 2019-a). Engaging and practicing in the data-centered world requires some understanding of what it means to be an ethical statistician and an ethical data scientist.

Ethics is defined as "the moral principles that govern a person's behavior or the conducting of an activity". That definition seems straightforward, but there is a problem: the moral principles that govern *a person* might not be the same ones that govern *me*. Two refinements on this definition may help us: "normative ethics" is involved with what most people (in a given context) would consider to be "right" or "wrong" ways of behaving. The fact that "most people" feel something is right or wrong means that feeling (about rightness or wrongness) is what makes that characteristic (rightness/wrongness) a *norm*, or normal, general, typical behavior. Obviously, people who are new to a community where there are norms may find it difficult to deduce or infer what is right and wrong; and when there is no formal community at all, then there can be no norms. More specifically, since normative ethics concerns itself with describing behaviors as right or wrong, when new behaviors become part of (or even just available to) the community, there is a clear need to determine whether they are right or wrong.

---

[4] This chapter includes material that was originally published in Tractenberg RE. (2016). Why and How the ASA Ethical Guidelines should be integrated into every quantitative course. *Proceedings of the 2016 Joint Statistical Meetings, Chicago, IL.* Pp. 517-535; and in Tractenberg RE. (2020, February 19). Concordance of professional ethical practice standards for the domain of Data Science: A white paper. Published in the *Open Archive of the Social Sciences* (SocArXiv), 10.31235/osf.io/p7rj2

So far, we have a few problems: Firstly, “behaving morally” needs to be described such that all persons are, not just “a person” is, governed by the same principles. If this is not the case, then anyone can behave any way they prefer and call it “moral”. Secondly, if rightness and wrongness are determined by norms, then “most people” need to be involved in the decisions about what behaviors fall onto the “Right” and “wrong” columns. When behaviors are determined to be “so wrong” that the community decides they should be punished, these tend to become barred by law (rather than simply by custom or norms). However, as you know/can imagine, something has to be described very precisely in order for it to be clear that a person has committed a crime or broken a law. Ethics tend not to be described so clearly – they do not rise to the level of “laws”, even when they are true cultural or community norms. This brings us to the third problem: there needs to be fairly clear descriptions of both what is “right” and what is “wrong”; but, as it turns out, so many things are “right”, and typically so few are “wrong”, it tends to be easier to focus on what is “wrong” – so we can educate all newcomers to a community about what to avoid. While that is certainly easier, it does not give the newcomer a sense of “how to behave” like someone who contributes to those “norms” of right behaviors. For these reasons, and many others, communities of professionals have drafted ethical guidelines, or codes of ethics, which describe what “the ethical practitioner” (or more precisely, *how* the ethical practitioner) does their job. The implication is that if a practitioner does not follow the guidelines/code, then they are not doing their job ethically.

There is another type or branch of ethics, “applied ethics”, which concerns itself with specific activities within well-defined communities (e.g., ethics applied to business is “business ethics”; ethics applied to biomedical sciences is “bioethics”). An important consideration for our purposes is, what if a quantitative practitioner sometimes works in business settings and other times in biomedical settings? Consultants may be described this way, because they are always quantitative practitioners, but the community in which they practice may change over time (this is also true for people who change jobs). Another situation is when a quantitative practitioner is actually working in *both* business *and* biomedical settings, like a biostatistician working for a pharmaceutical company. As you will see in later chapters, professional quantitatively oriented societies that have ethical guidelines (our example is the American Statistical Association, ASA) or a code of ethics (our example is the Association of Computing Machinery, ACM) state clearly that:

a. any practitioner who utilizes the tools/methods/techniques of the field are expected to follow *these* ethical guidelines/codes; and

b. any practitioner who is correctly following these guidelines/codes (i.e., because they are utilizing tools/methods/techniques of the field) should not allow people following other guidelines – other ethical norms – to cause them to *violate* the quantitative practitioner guidelines/codes.

Thus, these quantitative communities seek to ensure that their norms are followed whenever their tools/methods/techniques are employed. This is one way that these organizations – and communities – seek to promote *ethical* quantitative practice. These codes/guidelines are therefore standards for identifying "ethical practice" – as well as doing it, i.e., they are ethical practice standards. They outline "how to behave" like someone who contributes to "norms" of right behaviors in quantitative practices, and in general, when an individual fails to follow these norms, they are acting "wrongly". It is important to point out that "following the norms" should be observable – any two observers of the person acting should be able to agree, generally, on whether that was "right according to norms" or "wrong according to norms". If "acting wrongly" cannot be recognized sufficiently well to describe it to others (e.g., "I will know wrong behavior when I see it, but cannot describe it to you"), then it may not actually be sufficiently well-defined for everyone in the community to agree that it is, in fact, "wrong". This is another reason why guidelines are so important: only what is agreed on as "right" (or "wrong") are included in the guidelines, so people know fairly specifically what behaviors to avoid, and what behaviors mark the "ethical professional practitioner" (see Simonite 2018 for discussion of the relevance for this perception for "data science").

"Graduate instruction in statistics requires the presentation of general frameworks and how to reason from these." (Hubert & Wainer, 2011:62). This statement summarizes the perspective that teaching those who will be using statistics at work (graduate students across disciplines in their example) must acknowledge that the learners need to know more than just the formulas or how to use software. The learners also need what is sometimes referred to as "statistical literacy", or "statistical thinking", so that they can always identify the correct method, which exists within a general framework (e.g., categorical vs. continuous variables), and can also reason from the frameworks to the exact methodological features; and then reason about the results. The point Hubert & Wainer make is that "instruction in statistics" is not limited to just the statistical methodology: the frameworks *and* the ability to reason using those frameworks must both be learned. This also true for "instruction in ethics for quantitative practice": it requires the presentation of general *ethical* frameworks, and instruction and practice in how to reason from *these*. Professional societies have articulated guidelines that reflect the mindset of expert quantitative practitioners relating to their professional practice. These guidelines and codes are frameworks for ethical practice in the quantitative domains (Tractenberg, 2020).

Following the logic of Hubert & Wainer, though, providing the ethical guidelines or code of ethics to students or those new to the quantitative professions is not sufficient: "how to reason from these" is also a necessary component of ethics education. If you consider the Hubert & Wainer quote in the context of training for ethical quantitative practice, and you take the frameworks to be those professional practice standards, then "how to reason from these" is clearly the missing piece. "…(E)thics is not a vaccine that can be administered in one dose and have long lasting effects no matter how often, or in what conditions, the subject is exposed to the disease agent" (National Academy of Engineering, 2008 p. 36). Together, both the ethical frameworks *and* the ability to reason from these can shift even a single course in ethical practice from "one dose" that the National Academy of Engineering and National Research Council note is insufficient, towards the ability to reason ethically throughout a career, which is what the practice standards exist to promote.

This book was written to help familiarize the reader with these frameworks while also providing opportunities to practice reasoning from them. Some statements about ethical practice tend to –erroneously- reinforce the idea that ethical practice habits are simple to form (see Tractenberg, 2018). If this were true, it would mean that all of the unforeseen ethical challenges throughout statistics and data science were created purposefully and intentionally, because acting ethically is so simple and natural (that acting against those simple/natural instincts would require a lot of work on your part; Thiel 2015). While that might be a fair assumption for *illegal* behavior, it is not a reasonable conclusion about unethical behaviors.

Ethical challenges can arise in new and wholly unexpected situations throughout a career. Without training in how to reason with, or use, the ethical framework to prepare users of statistics for ongoing development of abilities to identify and reason through ethical challenges, it is not plausible to assume that these individuals will somehow prepare themselves. When "ethics training" focuses on static information or rules, the actual utility of that training is intrinsically limited. That is why this book focuses on ethical reasoning instead.

Every day and around the world, many individuals without professional statistician accreditation (PStat®) or even comprehensive training in statistics are asked to carry out statistical and data science tasks in both business and research settings –and this is an increasingly common situation as software, applications, and a perceived need for data analysis become ever more ubiquitous. It is untenable to assume that the training and practice working with ethical guidelines that are essential for developing the requisite familiarity with the GLs to support ethical statistical practice would be accomplished or even initiated by the single, *general-institutional* training module in "responsible conduct of research" that universities in the United States are required to provide for individuals receiving federal funding. This may be even less reasonable to assume for the institutional "ethics" training that many businesses and companies require of employees. Instead, if the ASA Ethical Guidelines and/or ACM Code of Ethics were introduced early, and reinforced throughout a curriculum to promote a sense of their relevance and ongoing engagement, this would result in long-term ethics education that would be obviously and specifically relevant to both students *and* faculty in the discipline of statistics. Moreover, trainees/students who are learning statistics or data science from or for *other disciplinary perspectives* would also learn both how to engage in these same important conversations about the ethical dimensions of statistical research and practice – *and also that such conversations are important*. Introducing ethical reasoning with/about data is important for improving the reproducibility of science across disciplines (see e.g. Freedman 2010; Collins & Tabak, 2014; McNutt 2014). For statistics and data science (encompassing practitioners in the disciplines of statistics, data science, and their intersection), attention to the features and standards of ethical practice will enrich both the career and the profession. This book exists to help with this enrichment.

2. Practice standards = Ethical Guidelines/Code of Ethics = Frameworks

The ASA Ethical Guidelines for Statistical Practice (American Statistical Association, 2022) comprise 8 core Principles plus an Appendix for organizations and institutions, which entail a total of 72 specific elements. The full (2022) Guidelines appear in following chapters, so here we just explore the topics:

A. Professional Integrity & Accountability (12)
B. Integrity of data and methods (7)
C. Responsibilities to Stakeholders (8)
D. Responsibilities to research subjects, data subjects, or those directly affected by statistical practices, Data Subjects, or those directly affected by statistical practices (11)
E. Responsibilities to members of multidisciplinary teams (4)
F. Responsibilities to Fellow Statistical Practitioners and the Profession (5)

G. Responsibilities of Leaders, Supervisors, and Mentors in Statistical Practice (5)
H. Responsibilities regarding potential misconduct (8)
APPENDIX: Responsibilities of organizations/institutions (12)

Meanwhile, the ACM Code of Ethics, updated in 2018, has four core areas, with 2-9 elements in each (ACM, 2018):

1. General Moral Principles (7)
2. Professional Responsibilities (9)
3. Professional Leadership Principles (7)
4. Compliance with the Code (2)

The 2022 Guidelines (ASA) and 2018 Code of Ethics (ACM) are presented in full, and discussed, in Chapters 1.3-1.5. These guidelines/codes are "the frameworks" that this book will help the reader learn to "reason from". At this point it is sufficient to notice the complexity and richness of these guidance documents. While certainly not the only such guidance, they are the professional standards for quantitative practitioners to be explored in this book. Because this book will teach the reader how to reason from *any such framework*, readers can go on to apply the reasoning to be presented to any other practice standard that is relevant (however, the practitioner should keep in mind that the ASA and ACM seek to promote ethical practice, and those communities hope that **their norms are followed whenever their tools/methods/techniques are employed**, in order to promote *ethical* quantitative practice). Other guidance may not be as comprehensive (e.g., International Statistics Institute, ISI, https://isi-web.org/index.php/activities/professional-ethics/isi-declaration has 12 principles[5]), but lack of a statement about something covered by ASA or ACM does not undermine the relevance or applicability of the ASA or ACM norms. This is primarily because these two organizations are among the largest – and therefore, most representative – for quantitative practitioners who are primarily statistical (ASA) or computational (ACM). The ISI ethical statement includes something that the ACM did not have in their 2018 version:

11. Bearing Responsibility for the Integrity of the Discipline
Statisticians are subject to the general moral rules of scientific and scholarly conduct: they should not deceive or knowingly misrepresent or attempt to

[5] International Statistical Institute (2010). Declaration on Professional Ethics (revised). Downloaded from http://www.isi-web.org/about-isi/professional-ethics/43-about/about/296-declarationprofessionalethics-2010uk on 11 Dec 2013.

prevent reporting of misconduct or obstruct the scientific/scholarly research of others.

The 2022 ASA Ethical Guidelines include several specific items that capture the same ideas (e.g., A1, B2-B3, C2, C8, D11, E4, H4-H5, H8). However, not all statistical practitioners, and data scientists, are doing scientific or scholarly work. Ethical reasoning, the 2018 ACM Code of Ethics, and the 2022 ASA Ethical Guidelines, are all specifically formulated for all practitioners. The guidance supports professional as well as scholarly and scientific applications of statistics, computing, and data science by practitioners at all levels.

Statisticians and data scientists, among other quantitative practitioners, have a special obligation to acknowledge and accept their responsibility for the integrity of the domain. In fact, Golde & Walker (2006) assert that, "Upon entry into practice, all professionals assume at least a tacit responsibility for the quality and integrity of their own work and that of colleagues. They also take on a responsibility to the larger public for the standards of practice associated with the profession." (p.10). This quote articulates an *assumption* about professional practice – namely, that everyone who enters a profession has been sufficiently prepared to take on the responsibilities for the standards that define that profession or discipline. The assertion was made (in the 2006 book on doctoral level preparation) in the context where individuals will have been formally trained as *stewards* of a particular discipline – those to whom the vigor, quality, and integrity of a particular discipline can be entrusted (Golde & Walker 2006: p. 5; see also Rios et al. 2019). The ASA and ACM do not (currently, in 2022) explicitly refer to disciplinary or professional stewardship, although the 2022 ASA Guidelines do specify (Principle F) that practitioners have responsibilities to both other practitioners and to the profession itself. When quantitative practitioners do not follow ISI statement 11, or the general practice standards of the ASA or ACM, then all those who make decisions on the basis of the results of quantitative practice may find their decisions, or their scientific or scholarly work, undermined.

Tractenberg (2016-a; 2016-b) discusses why training with the ASA Ethical Guidelines is important for training graduate and undergraduates in the quantitative sciences – these reasons pertain to the ACM Code of Ethics as well (see also Tractenberg, 2020). Four objectives or rationales for integrating training in ethical practice are to:

A. **Encourage ethical conduct in (throughout) the practice of statistics and data science**, by pointing out how everyone on a team has their specific role with its attendant obligations and priorities (i.e., to bring ISI principle 11 to *all* quantitative practice).

B. **Promote professionalism for all of the team members**, irrespective of their level of training in statistics and/or data science.
C. **Promote the consideration, prior to the start of analyses, of the analyses and the qualifications of the analyst/data scientist** to plan, execute, and interpret them.
D. **Engage with principles of professional practice for statisticians and data scientists**, which can promote both appreciation for the statistician and data scientist as a collaborating team member and understanding how this team member is accountable and responsible for their work.

Professional statisticians and data scientists are responsible for being ethical in their practice. However, responsibility for ethical behavior goes beyond the individual practitioner: both the ASA and ACM state that their practice standards are intended to promote ethical behaviors for both those who self-identify as their specific target audience (statisticians and ACM members) but are also applicable for *all those who utilize their technology/methods and approaches*. Further, both ASA and ACM practice standards have specific sections that address the responsibilities that employers (ASA) or leaders (ACM) have. It is impossible to require or even suggest that the statistician or data scientist is responsible for ensuring either that their attention to ethical practice is shared/honored by the rest of the team/their colleagues or that all participants in a project or workplace follow the ASA/ACM practice standards. However, when a practitioner encounters resistance or antagonism as they endeavor to practice ethically, it is evidence of violations of ACM Principle 3 (promoting ethical leadership in the workplace) and ASA Principle G (leading statistical practitioners so as to promote ethical behavior and an ethical workplace in which the statistician or data scientist feels comfortable).

It may seem like an unfair burden, but in fact, doing what you can to promote an ethical workplace strengthens not only your own practice but also that of your colleagues. This is not actually an expectation unique to statistical and quantitative practice; as noted, Golde & Walker assume this of all those who enter any professional context! However, statistics and data science are perhaps unique in the specificity of their ethical practice standards – which allows quantitative practitioners in these fields to actually follow the standards, and meet this expectation. One of the best ways to encourage real and positive attention to ethical practice in the workplace is to refuse to work in environments where ethical practice is not a priority or where it is actively resisted or discouraged. In some cases, the workflow is so distributed that there appears to be no single individual with the specific responsibility for ethical treatment of data, its analysis, or how the results are reported, maintained, and evaluated to ensure that harms do not accrue. Knowing what you know (or