# Chapter 3

## A Framework for Ethical Considerations

Data-empowered algorithms are reshaping our professional, personal, and political realities, and they are likely to have an even larger effect going forward. However, as with all developing technologies, increases in impact inevitably give rise to unanticipated consequences. These challenge our norms for how we use technology in ways consistent with our values. Many scholars, educators, and technology companies refer to these as **ethical challenges**, building on the applied ethics tradition from basic sciences.

Some challenges are best met by inventing improved or more nuanced technological approaches. However, many challenges will still arise based on how we deploy technology as products, or how statistical analysis interpretations guide law and policy.

While the word *ethics* may imply a branch of somewhat obscure philosophy, the applied ethical tradition is about both defining ethics and designing ethical processes clearly enough to help guide good choices. In the case of data science, it is also to develop programs that make good choices.

### 3.1 Professional Ethics

Companies and professional societies, including the American Statistical Association (ASA), the Institute for Operations Research and the Management Sciences (INFORMS), the IEEE, the Association for Computing Machinery (ACM), and Engineers Canada have long had important and useful ethical codes addressing matters of personal conduct and technical execution (American Statistical Association, n.d.; Association for Computing Machinery, 2018; Engineers Yukon, n.d.; Institute for Operations Research and the Management Sciences, n.d.). These include principles such as honesty, impartiality, and integrity.

The introduction to the ASA's code (American Statistical Association, n.d.) observes that:

> The discipline of statistics links the capacity to observe with the ability to gather evidence and make decisions, providing a foundation for building a more informed society. Because



https://doi.org/10.1017/9781009272230.005 Published online by Cambridge University Press

society depends on informed judgments supported by statistical methods, all practitioners of statistics – regardless of training and occupation or job title – have an obligation to work in a professional, competent, respectful, and ethical manner.

As the impact of statistics, operations research, and computing (and analogously, data science) has grown, many of these codes are being generalized to include broader societal considerations. Gotterbarn and Wolf write in a preamble to the 2018 ACM Code of Ethics that: "we find ourselves in situations where our work can affect the lives and livelihoods of people in ways that may not be intended, or even be predictable. This brings a host of complex ethical considerations into play" (Gotterbarn et al., 2018).

## 3.2 The Belmont Commission

In the human subjects research community, the *Belmont Report* is the central document of applied ethics in biomedical and behavioral research. In it, ethics is defined in terms of general principles (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1978). The Belmont Commission met monthly for four years in response to the 1932–72 US Public Health Service Syphilis Study at Tuskegee, a morally and scientifically flawed medical experiment. By including commissioners from a wide range of fields, including researchers, lawyers, administrators, and philosophers, the organizers hoped to protect human subjects while balancing societal norms, legal constraints, and society's need for innovation.

Despite its roots in the human subjects research context, the report outlines principles that are sufficiently general to be a basis for a useful ethical framework for data science research and products. In Belmont, these principles are called "respect for persons, beneficence, and justice." In more detail, they were then framed as follows:

- **Respect for persons.** This means ensuring the freedom of individuals to act autonomously based on their own considered deliberation and judgments. Often summarized as informed consent, this principle also includes having sufficient transparency to make judgments and also defending the autonomy of those with diminished consent, e.g., children or those who may be coerced into making a decision.
- **Beneficence.** Belmont encourages researchers *not* to limit their thinking to "do no harm," but to maximize benefits and balance them against risks. Doing so requires careful consideration of the immediate risks and benefits as well as a commitment to monitor and mitigate harms as results occur.
- **Justice.** The consideration of how risks and benefits are distributed, including the notion of a fair distribution. Fair may not mean "equal" but rather that the risks

are borne by the populations who stand to benefit (and are not borne by populations who will not ultimately have access to the fruits of the research).

These principles are intended to be broad and therefore applicable to yet unenvisioned technology changes and their consequences. At the same time, they are intended to be sufficiently specific that communities can come to a shared, deliberative consensus as to their implied best actions. In other words, from general, common principles, a community derives more context-specific standards and instance-specific rules. For a technologist, these rules imply even more specific design choices in modeling or in data product development.

This principled approach to ethics does not offer a single all-encompassing checklist that one consults for an answer that is the same in all contexts. Instead, principles are, by design, in tension with each other. They provide a basis to ask specific questions, which often do not have a right or wrong answer, but illuminate the tension in a situation or between positions.

## 3.3 Belmont Application to Data Science

The breadth of data science's impact argues for applying Belmont-like principles to it. Numerous scholars (Fiesler et al., 2015), researchers, and technologists have suggested how these principles can guide applied ethics even in the context of data-empowered algorithms. They sometimes also argue for extending the principles to emphasize the impact on society at large (US Department of Homeland Security, 2012; Vitak et al., 2016). However, as Belmont frames them, the principles provide a common vocabulary for researchers, data scientists, product developers, and regulators with which to reach consensus.

As co-author Jeannette writes in her 2020 essay, *Ten Research Challenge Areas in Data Science* (Wing, 2020):

> The ethical principle of Respect for Persons suggests that people should always be informed when they are talking with a chatbot. The ethical principle of Beneficence requires a risk/benefit analysis on the decision a self-driving car makes on whom not to harm. The ethical principle of justice requires us to ensure the fairness of risk assessment tools in the court system and automated decision systems used in hiring.

Another voice in this area comes from the European Commission, the executive branch of the European Union. Their *Ethics Guidelines for Trustworthy AI* shows that the need for ethical frameworks in technological areas is recognized by world governments, as well as researchers and ethicists (European Commission High-Level Expert Group on AI, 2019).

We, of course, accept there are distinctions between data-science-oriented implications of Belmont and human subject research, in particular medical research. In

https://doi.org/10.1017/9781009272230.005 Published online by Cambridge University Press

medical research, the principles motivate standards such as **informed consent** (a process of disclosing risks and benefits to an individual before gaining approval) and **fair selection of subjects**.

However, the digital domain can have different standards that are also consistent with the Belmont Principles, often because algorithms initiate automated actions. Here are some example considerations:

- Informed consent is hard to achieve in our current digital environment. To use a digital product, users must click "I agree," most often without comprehending the long and complex terms of service authorizing software actions over an extended period. Barocas and Nissenbaum identify "complex data flows" such as in digital services as possessing what they term a **transparency paradox** (Barocas & Nissenbaum, 2014). They argue that information disclosure provided to users is so simple as to be incomplete or deceptive, or it is so complex as to be incomprehensible. Data scientists adhere best to informed consent by respecting user norms at a level of transparency that avoids deception or unfairness, while allowing more detailed auditing and critique (e.g., via appropriate technical documentation or open source).
- For software, the risk/benefit balance of beneficence includes thinking through unintended consequences. It also requires the humility to recognize how hard it is to anticipate all the ways people will experience or use a product. That requires a commitment to monitor and mitigate harms as they are revealed.
- Justice, in the context of data-driven products, includes ongoing assessment of their fairness (technical and otherwise) as well as their training datasets. Justice includes fairness, with the understanding that defining "fair," even in technical communities, can be subjective (Narayanan, 2018) and is not as simple as giving it the same meaning as "equal." Our norms of justice also include an understanding of addressing and redressing prior harms, where possible. We say much more on fairness in Section 12.3.

We use the Belmont Principles to organize the analysis of several case studies in Part II. We then address the challenges to aligning ethics with a university's or technology company's data scientists' operational process in Part III. Part IV contains recommendations on how to proceed in the future.

While the Belmont Principles are our ethical starting point, we recognize that applications of data science may also require the consideration of other ethical principles upon which societies are based. For example, the principles of justice of war (jus ad bellum) and the conduct of war (jus in bello) are relevant to data science applications in military domains (Moseley, n.d.). Furthermore, data science is a sufficiently new field that we may eventually need to identify new relevant principles for its ethics.