

Reflections on “Ethics”

by

Robert McGinn
CCC Lead Ethics Investigator
April 4, 2023

I. What Is “Ethics”?

Asking “What is ‘ethics’?” to native speakers of U.S. English elicits diverse answers. A major reason why is that in English “ethics” has multiple meanings. Consider the following:

1. “Our ethics overlap.” Here, “ethics” is the plural of the singular noun “ethic,” where an “ethic” means a **prescriptive ideational framework** or system espoused by adherents. An ethic’s elements include specific ideas, principles, values, ideals, precepts, and rules, which adherents draw on in passing judgment on actions.
2. “I’ve been in ethics for years.” Here, “ethics” is a singular noun that means a **field of endeavor** in which some work or are otherwise active. Like other fields, ethics has a domain of phenomena which most in the field deem it legitimate to assess.
3. “‘Ethicists’ are practitioners of ethics.” Here, “ethics” means **systematic inquiry** into the acceptability of contested human actions. This sense is what is meant when someone is encouraged “to do ethics,” i.e., to investigate the acceptability of some contested action or practice.
4. “Ethics seeks to influence human behavior through judgments about the acceptability of actions.” Here, “ethics” refers to the general **cultural practice** of making and disputing prescriptive judgments about contested actions. Like the general cultural practice of law, the general cultural practice of ethics is used to encourage and deter certain behaviors. Ethics does so by relying on prescriptive judgments made by people with intellectual or spiritual authority and buttressed by their supporters, whereas law relies on the prospect of police power being invoked.

These meanings can be related. Drawing on their respective ethics (#1), people in ethics (#2) carry out ethics (#3) inquiries into actions, reach normative conclusions, and promulgate prescriptive judgements about them, thereby partaking in the general cultural practice of ethics (#4).

Put differently, “ethics” refers either to prescriptive ideational frameworks, or to a particular field of study, or to systematic inquiry into and evaluation of phenomena within the field’s domain, or to the general cultural practice of reaching, issuing, and contesting prescriptive judgments about disputed actions. Not surprisingly, failure to make explicit what one means by “ethics’ invites misunderstanding among those who disagree about the acceptability of an action.

II. Qualified Harm and Well-Being Consequentialism

Suppose a contested human action is to be examined with a view to ascertaining its ethical acceptability. What **criterion** should one use in such an inquiry? My preference follows:

the action's or practice's actual or most likely **consequences** for the **well-being** of affected and likely-to-be-affected parties.

The acceptability of a human action or practice hinges primarily on its consequences. The species of consequentialism I prefer is **qualified harm-and-well-being consequentialism**.¹ The acceptability of a work-related action of a scientist (S) or engineer (E) depends on the extent to which its consequences undermine or enhance the well-being of the affected parties.^{2, 3, 4, 5}

If a S's or E's work-related action or practice has no effects on, or seems unlikely to have any effects on, the well-being of any affected or likely-to-be-affected party, then talk of conducting an ethics inquiry into its acceptability makes no sense.⁶ Launching an ethics inquiry, exploring an ethical issue, or making an ethical judgment about an agent's action makes sense only if it has some actual or potential effect on the well-being of at least some affected parties.⁷

It should be noted that in determining an action's or practice's ethical acceptability, some ethicists use criteria that do **not** involve its consequences for affected parties. Here are four:

1. whether the action or practice has some *inherent property* that, *ipso facto*, makes it acceptable or unacceptable. For some, telling the truth, deliberately breaking a promise, or fulfilling a duty to serve one's country are actions that are right or wrong just because each (allegedly) has some inherent (positive or negative) property.
2. whether the action is *approved/disapproved* by some authority the inquirer accepts as supreme, e.g., "God," "The Party," or some religious or political figure.
3. what the *intention* is with which an action/practice is undertaken or carried out.⁸

¹ The modifier "qualified" signals that my adherence to "harm-and-well-being consequentialism" is conditional, as discussed on p. 6.

² Some ethicists limit the "parties" to be considered in ethical inquiry to human beings. Others construe "parties" more widely, encompassing "other sentient beings."

³ An action or practice may harm and benefit the same party, or harm one party or group while benefitting another. These complications are discussed below.

⁴ This is true whether the acceptability judgment is prospective or retrospective.

⁵ The extent to which an agent's action undermines or enhances the well-being of an affected party depends on the extent to which its consequences violate, leave unaffected, or fulfill the party's basic needs. For a classic but murky account of "basic human needs," see Abraham Maslow, *Motivation and Personality* (Harper and Row, New York, 2nd ed., 1970). For more philosophical discussion of human needs, see David Braybrooke, *Meeting Needs* (Princeton University Press, Princeton, 1987).

⁶ Except to those who hold that the acceptability/unacceptability of actions is independent of their consequences.

⁷ Significant negative effects of an action or practice on a party's well-being count as harming that party.

⁸ Although the intention with which an action is undertaken by an agent is irrelevant to its acceptability – the road to hell/heaven can be paved with good/bad intentions – the agent's intention is relevant to judging her or his character.

4. whether the action or practice is what the inquirer's *model of virtue* would do or would have done under the prevailing circumstances.⁹

III. Fundamental Ethical Responsibilities of Scientists and Engineers (FERSEs)

As a qualified harm-and-well-being consequentialist, when appraising a S's or E's action, I've found it useful to take **an ethical responsibilities approach**. This involves looking far and wide to discern all ethical responsibilities applicable to the situation at hand.¹⁰

Key to determining what ethical responsibilities a S or E has in a particular work situation is the fact that technical professionals who function in society and who are employees or have clients have four basic, harm-rooted, work-related ethical responsibilities: **the Fundamental Ethical Responsibilities of Scientists and Engineers (FERSEs)**.

Here are *rough* formulations of the FERSEs of employed Ss and Es:

FERSE1: to not **cause** harm to others through her/his work.

FERSE2: to **prevent** harm to others from occurring through her/his work.

FERSE3: to **alert** parties at risk of harm from her/his work.

FERSE4: to **serve** the interests of her/his employer or client.

While concise, these formulations are imprecise and can be misleading. Less concise but more precise formulations follow:

Any S or E has fundamental ethical responsibilities....

- to not cause or contribute to causing harm, and to not create or contribute to creating an unreasonable risk of harm, to others (or to public welfare or the public interest) through her/his work (**FERSE1**).
- to try to prevent harm or an unreasonable risk of harm to others (or to public welfare or the public interest) from her/his work, from work of others in which s/he is involved, or from work of others with which s/he is familiar and about which s/he is technically knowledgeable (**FERSE2**).
- to try to alert and inform individuals and segments of the public at significant risk of being harmed that they are vulnerable to that risk from her/his work, work of others in which s/he is involved, or work of others with which s/he is familiar and about which s/he is technically knowledgeable (**FERSE3**).

In addition, any S or E who is employed -- by a company, government, or other type of organization -- or who works for a client has a fourth fundamental ethical responsibility:

⁹ There are good reasons for not adopting any of these four alternatives as the criterion to use in appraising acceptability. However, those reasons will not be discussed here.

¹⁰ One limitation of this approach is that it is more helpful in getting clear about the relationship of the S's or E's conduct to affected-party harm than it is about its relationship to affected-party well-being.

- to work to the best of her/his ability to serve the legitimate interests of her/his employer or client (**FERSE4**).¹¹

IV. Further Clarifications of the FERSEs

Even the more precise formulations of the FERSEs need further clarification:

- A. FERSE1 can be violated not only by acts of *commission* intended to harm others, but also – arguably more frequently -- by (negligent) acts of *omission* that unwittingly contribute to causing harm or to creating an unreasonable risk of harm to others, to public welfare, or to the public interest.
- B. The word “try” in FERSE2 and FERSE3 is critical. FERSE2 and FERSE3 do **not** oblige Ss or Es to *succeed* in preventing harm or to succeed in alerting and informing about the risk of harm. Sometimes doing so is impossible, and one cannot have an ethical responsibility to do that which is impossible. What FERSE2 and FERSE3 oblige the S or E to do is to **try** to prevent and to **try** to alert and inform.
- C. The word “legitimate” in FERSE4 is also critical. The S or E employee does not have a fundamental ethical responsibility to do her/his best to serve **all** claimed interests of the employer or client; only those that are ethically **legitimate**.¹²
- D. FERSE4’s ethical responsibility to work to the best of her/his ability to serve the legitimate interests of the employer or client is incumbent on the S or E *only as long as the employer or client treats her/him fairly and reasonably re compensation and the conditions of work*. Should this condition cease to be met, the S or E is no longer bound by FERSE4.
- E. While the specific ethical responsibilities of a S or E in a particular work situation depend on the harm-centric FERSEs, they are not always the final word on what s/he should do in that situation. Rather, they are *general guides* to what the S or E should do, ones that *always* merit serious consideration and are *initially binding*. Those who believe that the S or E should act differently in a particular situation than the FERSEs and the features of the situation suggest bear the burden of making a compelling case to that effect.

For example, given FERSE2, a S or E might have a *prima facie* ethical responsibility in certain kinds of situation to blow the whistle publicly. However, if it could be shown that doing so in a specific situation would jeopardize national security while at best preventing only relatively minor harm, that consideration might trump the *prima facie* ethical responsibility and make **not** blowing the whistle publicly the ethically right course of action, all things considered, in that situation.

¹¹ FERSE4 is a more defensible version of the traditional employee-loyalty-to-employer precepts found in many codes of engineering ethics. For example, Fundamental Canon I.4 of *The NSPE Code of Ethics for Engineers* states, “Engineers, in the fulfillment of their professional duties, shall act for each employer or client as faithful agents or trustees.”

¹² Ethically illegitimate employer or client interests include stealing a competitor’s intellectual property, bringing a risky product to market without adequate safety testing, and cheating in satisfying applicable government regulations that affect when a new company product can be released to market.

- F. Suppose that, under FERSE4, an employed S or E has an ethical responsibility to carry out a certain course of action to advance a legitimate employer interest, but also, under FERSE1, an ethical responsibility to *not* carry out that same course of action, e.g., because of the harm or unreasonable risk of harm to others it would cause or create. How should the S or E faced with such a conflict proceed? When, in a specific situation, FERSE1 and FERSE4 conflict, the S or E should presume that FERSE1 takes precedence over FERSE4. (This is so because repeated failure by Ss or Es to give priority to not causing public harm over serving employer or client interests could result in society seriously restricting the autonomy of Ss and Es, something that would arguably significantly harm society.) However, the S or E should also be open to the possibility that in a specific situation a compelling case might be made that overcomes that presumption, justifying giving FERSE4 precedence over FERSE1 and the S or E acting accordingly.¹³
- G. Suppose a S or E is contemplating an action which, while it would harm some parties, would also benefit them and/or different parties. Such a situation is fairly common. How should the would-be ethically responsible S or E approach such a situation? Some contend that the only thing that matters is *whether the contemplated action's benefit exceeds its cost*. But financial cost is not the only harm relevant to ethical inquiry. Hence, perhaps the S or E needs only to ascertain *whether the action is likely to produce more benefit than harm* and, if so, act accordingly.

However, contrary to widespread belief, the fact that an action or practice has a surplus of benefit over harm, i.e., that its expected aggregate benefit exceeds its expected aggregate harm, is *not always sufficient* to make it ethically proper (all things considered) for the agent to proceed. For there are conditions – I call them “trumping conditions” – that, if any applies to the situation at hand, make it reasonable or obligatory to **not** proceed, *even if* the expected benefit exceeds the expected harm.

Here are three such conditions: (1) the benefits and harms are likely to be unjustly distributed, e.g., the currently-worst-off or some relatively powerless group will bear (or is highly likely to bear) the bulk of the resultant harm or risk of harm, and/or receive the smallest share of the resultant benefit; (2) the magnitude of expected harm, although less than the expected benefit, exceeds some limit of ethical

¹³ Suppose an engineer, Smith, is an employee of the U.S. National Security Agency (NSA). Protecting U.S. national security is clearly a legitimate NSA interest. Suppose further that, pursuant to that interest, NSA instructs Smith to design a new software program that, when applied, would covertly gather personal medical and financial data about a large number of U.S. Internet and email users. Use of this work product would arguably cause significant harm by violating the privacy interests of those whose records were covertly captured. Put differently, Smith's design activity, while in accord with FERSE4, would conflict with FERSE1. Smith should approach her/his consideration of the situation with the presumption that not violating FERSE1 takes precedence over not violating FERSE4. However, s/he should also be open to the possibility that a compelling case might be able to be made that would override that presumption, and justify Smith's designing the requested software. Whether such a case could be made would depend on *the detailed specifics of the situation*, such as the magnitude and scope of the harm likely to be done to civilians through use of the software, the weight of the NSA's national security interest that prompted the assignment to develop the software, the magnitude and risk of the harm to the country that could result from not developing and using that software, and whether alternative courses of action exist that would further NSA's legitimate national security interest as much as developing and using the software at issue would, but cause significantly less harm to civilian informational privacy interests.

acceptability (e.g., assuming commensurability, while 50.5 units of benefit “exceeds” 49.5 units of harm, the 49.5 units of harm might reasonably be deemed too heavy a price to pay for realizing the marginally greater benefit); and (3) the expected harms are serious and practicably irreversible over time.

Only if no such trumping condition applies to the situation at hand would the fact that the expected benefit exceeds the expected harm be sufficient to justify proceeding with the action or practice in question. If any trumping condition applies, then acting in line with FERSE1 to avoid causing harm, even when that harm is exceeded by the expected benefit, might still be ethically right all things considered.

The more precise formulations of the **FERSEs** contain three critical terms: “**harm**,” “**risk**,” and “**cause**.” Brief remarks on each follow.

1. “**harm**”: when probing the acceptability of human conduct, using a broad notion of harm is imperative. That is, one must have and use a concept or idea of harm that encompasses *all* of its kinds; not just tangible physical and quantitative financial harms, but also psychosocial, institutional, cultural, and environmental harms. All harms, even those indirectly caused, to any and all affected parties, including individuals, groups of humans, other sentient beings, the public, and important social institutions, must be considered.
2. “**risk**”: Ss and Es involved in a project sometimes significantly *underestimate* the magnitude of the risk of harm to others that it poses. One reason that happens is that they assume that the risk posed depends only on the nature or design of the object or system they’re working on or proposing to implement. They sometimes fail to take into consideration the fact that the magnitude of the risk of harm their work poses is also a function of the **CONTEXT** in which the action is carried out and/or in which its product will be applied, used, and take effect. In particular, *ethically responsible risk assessments of technical endeavors, products, or systems designed or carried out by Ss or Es* must take into consideration **contingent contextual factors**, such as...
 - the cultures of the labs, manufacturing facilities, construction companies, and regulatory agencies in question
 - the training and caliber of the lab researchers, production and construction workers, and regulatory personnel
 - the rigor of the regulatory process
 - whether any of the firms doing the work are under great temporal or economic pressure
 - the quality of the materials and reliability of the equipment used
 - whether the Ss, Es, or other workers would be held personally accountable for anything that goes wrong and harms others

Taking such factors into account is crucial if a S or E wants to fulfill the second part of **FERSE1**: to not cause or contribute to causing an unreasonable risk of harm to others through one’s work. Ss and Es sometimes fail to ensure that their product or system risk assessments are *realistic*, e.g., by implicitly assuming that everything technical and social in the risk situation will work just as envisioned by the designers. This form of idealization can lead to over-optimistic risk assessments, hence to approval of unreasonably risky projects, hence to actions that violate FERSE1.

3. **“cause”**: the work of Ss and Es sometimes *directly and immediately* causes harm to some parties affected by it. Less obviously, even if such work is not the direct, sole, tangible, and immediate cause of whatever harm or unreasonable risk of harm results, it can be a **contributory causal factor** to its downstream occurrence. Granted, a non-technical political actor might decide to use a novel but risky engineering product or system developed by Ss and Es, thereby triggering/precipitating harm or unreasonable risk of harm to others. However, even so, it may still be that (i) the work of Ss and Es *enabled* or *set the stage for* such a decision to be made, (ii) the Ss and/or Es may have misrepresented to the non-technical decision maker the risk their new product or system poses to others, and (iii) the Ss and Es may have been aware that the lay decision-maker would in all likelihood decide to deploy or use the product or system in ways apt to cause significant harm or unreasonable risk of harm to others. In such cases, while the work of Ss and Es might not be the *triggering* cause of the harm in question, it might be a *contributory causal factor* in its occurrence. Contributory causal factors to the downstream occurrence of an outcome are of various sorts, including facilitating, enabling, stage-setting, incentivizing, and stimulating factors.

V. Using the FERSEs in Gauging the Ethical Acceptability of a S’s or E’s Action

Suppose a S or E in a work situation with specific personal, social, technical, political-economic, cultural, and environmental characteristics is trying to decide what, from an ethics perspective, s/he should do. Assuming s/he grasps the FERSEs, one option is to navigate a four-step process: (i) **identify the work situation’s important social and technical characteristics**; (ii) **identify the FERSEs that apply** to it; (iii) **bring the applicable FERSEs to bear on the situation** and **deduce the specific derivative ethical responsibilities** incumbent on her/him; and (iv) check whether the likely harm-and-well-being consequences of acting in accord with those responsibilities are problematic enough to justify overriding them and acting accordingly.

For example, suppose a S or E has reason to believe that her/his work (or the work of another S or E with which s/he is familiar and about which s/he is technically knowledgeable) poses a serious risk of causing (or of contributing to causing) substantial harm to certain parties likely to be affected by it, e.g., fellow workers or consumers. Further, suppose that most of the likely-to-be-affected parties are oblivious to the risks involved, have not been warned about them, have not been trained to avoid them, and the regulatory agencies charged with protecting workers and consumers are notorious for cursory oversight work. FERSE2 arguably applies to that situation. By applying FERSE2 to that situation, with its distinctive technical and social features, the S or E should be able to determine whether s/he has a presumptive derivative ethical responsibility to *publicly* blow the whistle in an effort to try to prevent (or mitigate) that unreasonable risk of harm, or to *publicly* alert and inform all vulnerable parties that they are vulnerable to harm in that situation. If either of those derivative ethical responsibilities applies, the S or E should assess whether acting in accord with it would inadvertently cause harm of such a nature – re its distribution, magnitude, or reversibility -- that the responsibility should be overridden and the S or E should refrain from publicly trying to prevent or alert in that situation.

To recap, instead of relying on “religion,” “feeling,” “intuition,” or “conscience,” and instead of doing what a work supervisor, close friend, colleague, or professional society code of ethics instructs her/him to do, the would-be ethically responsible S or E should consider utilizing the following framework of ideas:

1. the general FERSEs
2. the S's or E's ideas of harm and well-being
3. the full set of parties affected or likely to be affected by a contemplated action or practice in question
4. the concepts of cause of harm and contributory causal factor in the occurrence of harm
5. idealized vs. context-sensitive engineering risk assessments of new products or systems
6. the cultures of pertinent labs, construction sites, manufacturing facilities, and regulatory agencies involved with a particular product or system on which the S or E is working
7. contingent risk-related features of the multi-leveled techno-social context of the scientific or engineering work in question
8. the specific ethical responsibilities of the S and/or E derivable by applying the relevant general FERSE(s) to the specific technical work situation/context in question

If a S or E uses this framework, her/his thinking about what is ethically acceptable, right, or proper in a specific work situation should be clarified. Identifying the salient social and technical features of the concrete situation and the applicable FERSEs are important opening steps in determining what, from an ethics perspective, the S or E should do in a specific work situation.¹⁴ The derived, context-specific ethical responsibilities are *initial guides* to what the S/E should do in that situation. These guides *always* merit serious consideration. However, the guidance they provide is provisional. It is possible that, all things considered, including enhancements as well as violations of well-being and how both are likely to be distributed over the affected parties, the S or E should act differently than the derivative ethical responsibilities prescribe. But a compelling case to that effect would need to be made to override the initial guidance.

¹⁴ For specific ethically responsible and irresponsible practices of Ss and Es in research and innovation, including during the understudied entrepreneurial phase, see the author's "Beyond 'FFP': A Synoptic Matrix of Ethically Responsible and Irresponsible Practices of Scientists and Engineers in Research and Innovation."